Streetlight Halos

Walter Tape
University of Alaska Fairbanks

June 9, 2010

Contents

Preface

Halos form by the refraction or reflection of light in ice crystals in the atmosphere. Normally the sun or moon is the source of the light, in which case the light rays reaching the crystals are nearly parallel. Those halos I will call the classical halos. But in cold climates where ice crystals occur in the atmosphere near the ground, the source of the light can be a streetlight or other artificial light, and then the incoming light to the crystals is divergent. The divergence makes most such halos remarkably spatial. Any halo formed in divergent light I will call a streetlight halo or a spatial halo. Streetlight halos are the topic here.

What may appear here to be a book is really just a collection of articles, most of which I originally wrote and distributed to a few friends starting in January of 2009. When I recently gathered the articles together, I made a few changes, mostly for the sake of consistent terminology and notation, but I tried not to tamper too much with the original text, in hopes of preserving and conveying some of the fun and excitement that I had experienced initially. As a result, there is some repetition from one chapter to the next, and the chronology in general leaves something to be desired.

My treatment of streetlight halos here is rather one-sided. I mainly wanted to see whether I could analytically compute the spatial curves and surfaces that are the halos. So I have included very few photographs of real halo displays, and I have included no simulations of halos at all. The website *Ice Crystal Halos* [1] has photographs, and Gislén et al [6] have simulations.

Most of you are already hooked on streetlight halos. If so, you know that they can be dramatically spatial, almost psychedelic at times. Diamond-like sparkles delineate exotic curves and surfaces in space, flashing on and off as moving ice crystals pass through. Many of the halo figures in the following pages are beautiful, but they cannot match what you get in a real streetlight display.

I am assuming that you are familiar with the classical halos. If not, please see Ref. [10] or Ref. [2].

Chapter 1

Parhelic Circle Halos

Lately the website *Ice Crystal Halos* [1] has been loaded with great photos of streetlight halos. One photo in particular, at the lower left in Jukka Ruoskanen's collage posted January 6 and shown in Fig. 1.1 below, caught my eye; I had never seen anything like it. So I was prompted to think more about streetlight halos—by which I mean any halo formed in a divergent light source. The diagrams in this chapter show streetlight displays for a few special halos, all of which are on the parhelic, superparhelic, or subparhelic circles. They are halos whose spatial (i.e., streetlight) versions are relatively easy to compute, but they nevertheless give some nice insights into spatial halos in general. And one of them seems to solve the puzzle posed by Jukka's photo.

Figures 1.3 and 1.4 show the halos for a lamp elevation of ten degrees. You will recognize the superlamp and sublamp halos; they would normally be referred to as light pillars, but there are good conceptual reasons for avoiding the pillar terminology here. The parhelia will also be familiar if you have seen streetlight halo displays. And you will also recognize the superparhelia, if not from personal experience then at least from Fig. 1.2. Obviously these diagrams are highly idealized and not very realistic, and I am not going to try to give all of the fine print here.

Most of the fun in streetlight displays comes from their amazingly three-dimensional feel, but you don't get that feel from diagrams like Fig. 1.3 and Fig. 1.4. Figure 1.5 therefore gives a spatial version of Fig. 1.4. In fact, the plane figure (Fig. 1.4) was made from the spatial figure (Fig. 1.5) by projecting the colored curves and surfaces in the spatial figure to a plane, using the lens projection with camera placed at the observer's location. So the plane figure shows what the colored curves and surfaces in the spatial

Figure 1.1: Streetlight (spotlight, actually) halo display January 1-2, 2009, from Jukka Ruoskanen's website [3]. ©Jukka Ruoskanen.

Figure 1.2: Another of Ruoskanen's photos from January 1-2. ©Jukka Ruoskanen.

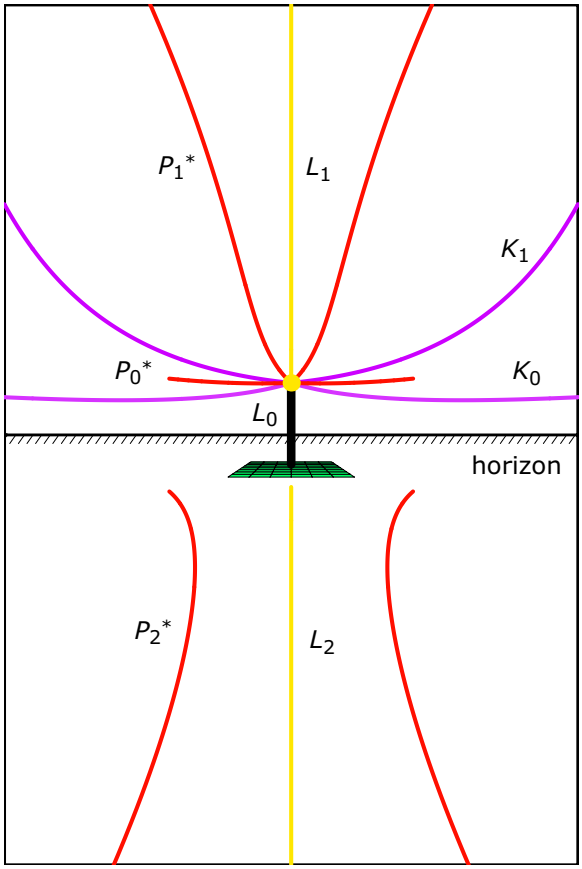

Figure 1.3: Idealized depiction of some halos in a streetlight. The angle of elevation of the lamp above the observer is $\Sigma = 10°$. Although the lamp base is shown, the halos are depicted as if there were no ground to interfere. The projection is for a 12 mm rectilinear lens. The halos are:
P_0^*, P_1^*, P_2^*: caustics of the parhelia, superparhelia, and subparhelia.
K_0 and K_1: 120° parhelia and 120° superparhelia.
L_0, L_1, L_2: lamp (halo), superlamp, and sublamp.

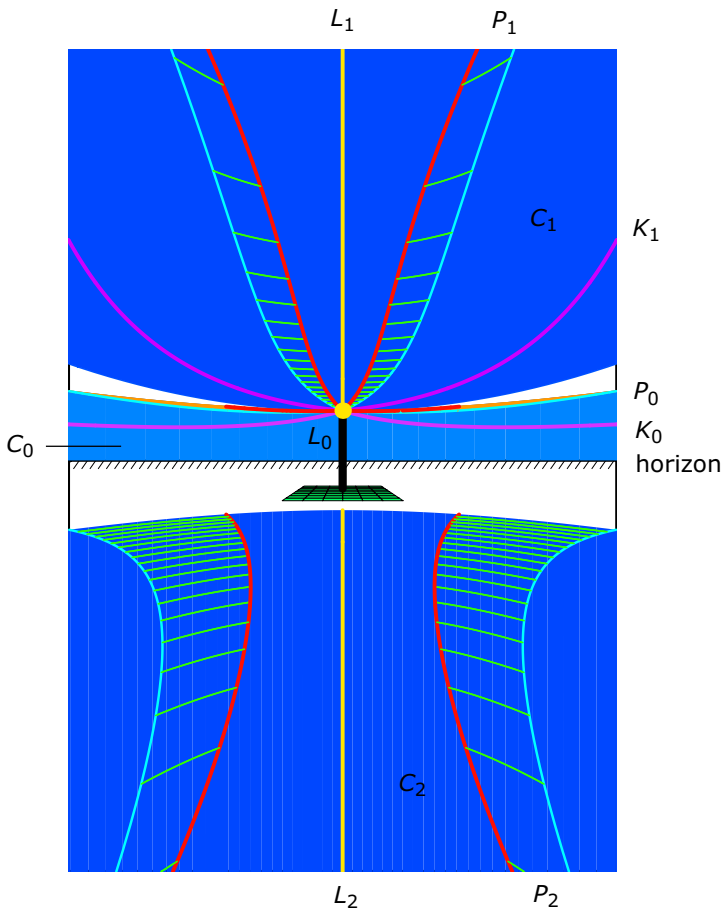

Figure 1.4: Same as Fig. 1.3 but with the addition of the parhelic, superparhelic, and subparhelic circles C_0, C_1, C_2 (all blue), and showing the complete parhelia, superparhelia, and subparhelia P_0, P_1, P_2 rather than just their caustics.

figure would like like to the observer if he (or she) had no depth perception.

In Fig. 1.5 the middle surface is the parhelic circle C_0, and the upper and lower surfaces are the super- and subparhelic circles C_1 and C_2, respectively. Each of the colored halos is on one of the three surfaces: the halos L_0, P_0, K_0 are on C_0; L_1, P_1, K_1 are on C_1; and L_2, P_2, K_2 are on C_2. The orange stippled regions with the red and turquoise boundary curves make up the parhelia P_0, and the green stippled regions, also with red and turquoise boundary curves, make up the super- and subparhelia P_1 and P_2. The red curves are the caustics or minimum deviation curves;[1] on those curves the ray paths through the halo making crystals are symmetric, as shown in Fig. 3.8 of Chapter 3. The turquoise curves are the maximum deviation curves; that is where the ray paths through the halo making crystals are tangent to the crystal at entry or exit.

People who have seen these halos in streetlights may agree that Fig. 1.5 correctly conveys some of the spatial feeling that the observer senses, but nobody will see halos as in the figure, viewed from the reader's perspective. That is, the figure shows the curves and surfaces in space where crystals can send light *to the observer* via the relevant ray paths. But those same curves and surfaces are not apt to be sending any special light in the direction of the reader.

In Fig. 1.5 the "lamp line" is the yellow line segment L_0 joining the observer and the true lamp. This is where the crystals are located that send light to the eye undeviated. Once you catch on to the meaning of the figure, you will agree that it makes sense to include the lamp line as a halo. Of course to the observer it is not a very interesting halo. He sees it only as a point—the true lamp. Nobody will ever see it in reality as a line.

1.1 What is a halo?

You can argue whether the curves and surfaces in Fig. 1.5 should be called halos. The man in-the-street—the guy who is always asking where the rainbow is located—would say yes. Most of us think we know better: "Halos are just collections of directions, or points on the celestial sphere. They are not located anywhere. Etc., etc." But the man-in-the street may be right. These curves and surfaces in space perhaps *are* the halos. Rather, they are the *potential* halos.[2] Each observer is carrying around his own set

[1] This coincidence of minimum deviation with caustic is a feature of the parhelia and a few other special halos; it does not work in general.

[2] A more sophisticated approach would be to replace each halo curve or surface by an intensity distribution in space; the distribution would be concentrated near the curve or

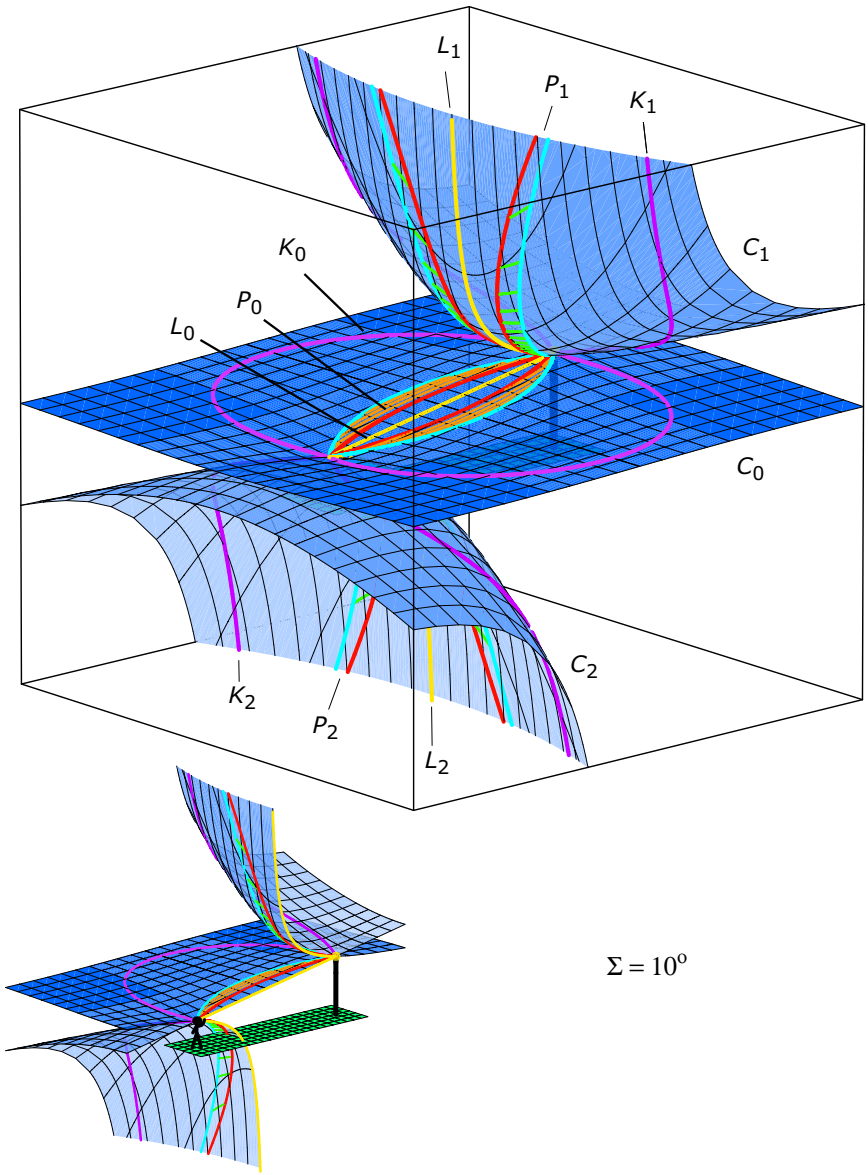

Figure 1.5: Spatial counterpart of Fig. 1.4. The figure shows the regions of space where crystals can send light to the observer via the relevant ray paths. The parhelic, superparhelic, and subparhelic circles C_0, C_1, C_2 are the three blue surfaces. The parhelia P_0 are the orange-stippled regions of C_0 with red and turquoise boundary curves. The super- and subparhelia P_1 and P_2 are the green-stippled regions of C_1 and C_2, also with red and turquoise boundary curves. The small diagram is the same but cut away to show the observer and the streetlight.

of potential halos—a collection of exotic curves and surfaces in space. A portion of a halo gets activated when appropriate crystals pass through that portion. To me this point of view has been helpful.

Figure 1.4 was made from Fig. 1.5 by assuming that all of the halo curves and surfaces were activated, that is, that all of space was filled with crystals and that light could proceed unimpeded from lamp to crystals to observer. You can learn a lot by examining the figure and thinking about which halos would disappear when the assumption fails in various ways, that is, when only parts of the halo curves and surfaces are activated. For example, the observer is normally standing on the ground, with no crystals below him. The subparhelic circle (surface) then cannot be activated, so the subparhelic circle and all of the halos on it must disappear. And a streetlight will often have a shade on it. If the shade is such that it blocks light rays having an upward component, then the superparhelic circle (surface) and its halos will disappear as well.

If your halos are forming in a spotlight instead of a streetlight, then the beam of the light activates the portions of the halo curves and surfaces that it shines on. Perhaps the light is shining directly at the observer, but it need not be so. It can point in any arbitrary direction, and you can anticipate the halo display seen by the observer by noting what portions of the halo curves and surfaces get activated by the beam.

Another very concrete example: At night in cold climates you sometimes see halos forming in ice crystals on your car windshield in light from a streetlight. To predict the halos using Fig. 1.5 you just add the windshield to the figure and take the intersections of the various halo curves and surfaces with the windshield, and throw away everything else; it is the windshield that is activating the halos. Many conditions would obviously need to be satisfied before this example would be realistic; I only mention it because it helps in thinking about the meaning of Fig. 1.5.

1.2 Eye-connected halos

Let's imagine one more way to activate the halo curves and surfaces. Suppose that there is only a small (but effective) spherical swarm of crystals surrounding the observer. Then it is that spherical region that activates the halos. To see what the observer perceives, you draw the sphere around the observer's position in Fig. 1.5 and throw away all of the halos outside the sphere. Or we could just zoom in on the observer, as in Fig. 1.7, and use only

surface.

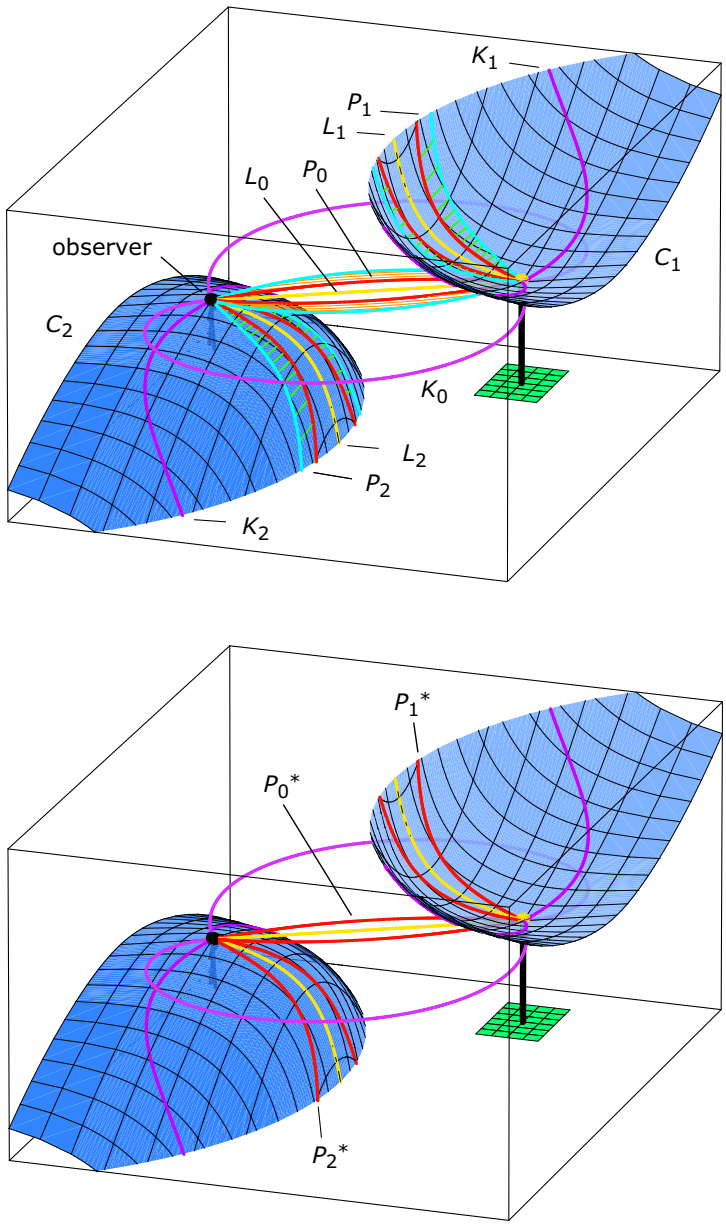

Figure 1.6: (*Top*) Another perspective on Fig. 1.5. The parhelic circle surface C_0 has been removed for clarity, leaving the parhelion surface P_0, the 120° parhelion curve K_0, and the lamp line L_0 all hanging in space. (*Bottom*) Same except that the tails of the parhelion, superparhelion, and subparhelion surfaces have been removed, leaving their caustic curves P_0^*, P_1^*, P_2^* (red).

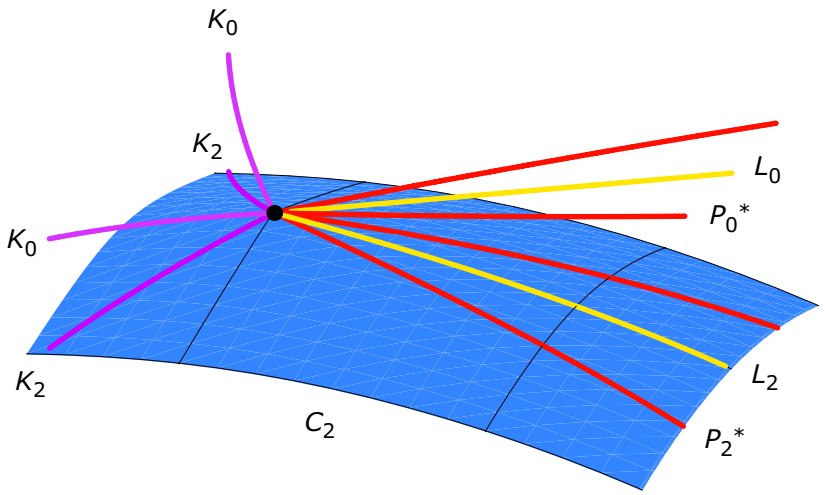

Figure 1.7: Halo curves from Fig. 1.5 that are eye-connected, that is, that reach the observer's eye. In this view, zoomed in on the observer's eye (black dot), the tails of the parhelia and subparhelia have been removed, leaving their caustics P_0^* and P_2^*. The parhelic circle surface has also been removed, in order to expose the halo curves on the subparhelic circle surface C_2. Each halo curve is seen to be nearly straight here, and each, leaving the eye, points in the direction of its classical instance. In a sense, this diagram *is* the classical instance of the halo display, since the only halo-making crystals are so close to the observer and since the light source is so far away.

those halo curves and surfaces in the zoomed-in view. In the figure you see fragments of the lamp line and the sublamp curve, the parhelion caustic and subparhelion caustic curves, and the 120° parhelion and 120° subparhelion curves. Each of those fragments is nearly straight (the lamp line fragment *is* straight) and so the observer perceives each as nearly a point, not a curve. All else being equal (a big caveat!), that point should be relatively bright, because in that direction there are lots of crystals—one behind the other, so to speak—that are sending light to the eye. The bright point is essentially the classical (that is, parallel, not divergent, light) instance of the halo in question. After all, it no different than if the lamp were the sun and if the little crystal swarm here were a cloud of crystals high in Earth's atmosphere. Thus the classical halos are what you get by activating the halo curves and surfaces on just a small sphere centered on the observer, as in Fig. 1.8. The classical halos are a special instance of the spatial halos.

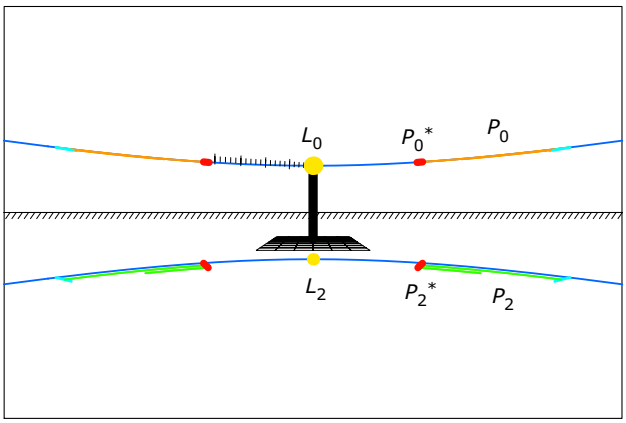

Figure 1.8: Similar to Fig. 1.4 but only using crystals that were closer to the observer than five percent of the horizontal distance between the observer and the lamp. The resulting halos approximate familiar classical halos—the parhelia and superparhelia with their caustics, and the subsun (and sun). The blue curves are the classical parhelic and subparhelic circles. The larger tick marks are at five degree intervals. The focal length for the projection is 15mm.

When a halo curve or surface reaches the eye, I will say that it is *eye-connected* (for the elevation Σ of the lamp). To the observer, the eye-connected halos in a streetlight display have a distinctive look. Let's remove the restriction on the crystal swarm, so that lamp light is again reaching crystals everywhere in space. In fancy language, all the halo curves and surfaces are being activated. Each eye-connected halo curve should still be bright in the direction where it leaves the eye, since, just as in the case of the small crystal swarm, the observer is getting lots of light from crystals in that direction. But now, unless the halo curve is itself straight (the lamp line), the observer perceives it as a curve that ends on the bright point. The brain tries to place that point at some distance away, and you do not get the impression that the halo curve reaches your eye. I see no reason why you *should* get that impression, but I do not know what determines the apparent distance to the point. In any event, each eye-connected halo curve appears to end at a point which, if projected to the celestial sphere, is the classical instance of the halo. This is illustrated in Fig. 1.9.

In practice the curve appears to terminate not on a perfect geometric point but on an ill-defined blob of light. The blob seems mysteriously suspended at some distance in the direction of the classical halo.

Each eye-connected surface halo consists of many curves reaching the eye, and it can be at least partly understood by thinking about its constituent curves as above. That is, each of those curves will appear to end in a diffuse little blob of light, and those blobs will fit together to form a perceived boundary curve of the surface halo, where the surface seems to come to an end. For the parhelic circle surface, for example, the curve in question would be the upper boundary of the parhelic circle C_0 in Fig. 1.4 or the upper blue curve in Fig. 1.9, but what does the curve look like in space? For the parhelic circle surface, we know that each blob will lie on the imaginary vertical cone that has vertex at the eye and that passes through the lamp. The curve of blobs would be like a rubber band going around the cone, hugging it. Projected to the celestial sphere from the observer's eye, the rubber band would appear as a circle, but the observer's brain may not opt for the projection. So at what distance will the brain place those blobs and hence the rubber band? If it puts them all at the height of the lamp, then the rubber band becomes a horizontal circle passing through the lamp. But I doubt that it works that way, in which case our parhelic circle boundary curve—the rubber band—might seem to drape around the cone in some exotic fashion, still clinging to it.

I, myself, have only seen a good streetlight parhelic circle once, long ago, and I did not pay enough attention to it. Marko Riikonen has seen at least

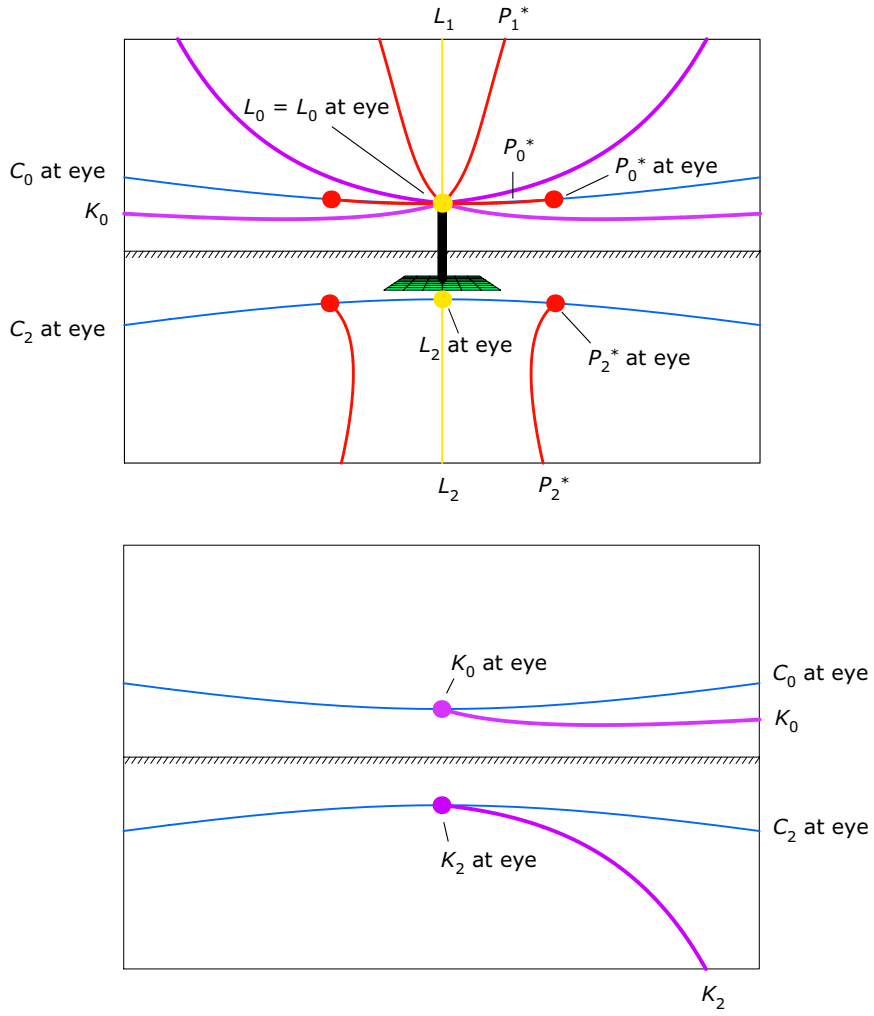

Figure 1.9: Illustrating that each eye-connected halo curve H appears to end at the point 'H at the eye' that is the classical instance of the halo. As in Fig. 1.7 the eye-connected halo curves here are the lamp and sublamp halos L_0 and L_2, the parhelion and subparhelion caustics P_0^* and P_2^*, and the 120° parhelia and subparhelia K_0 and K_2. (*Top*) Same as Fig. 1.3 but with a colored dot added at H-at-the-eye for each eye-connected halo curve H. So the four red dots are at the classical parhelion and subparhelion caustic points, and the two yellow dots are at the lamp and at the classical sublamp point. (*Bottom*) Same but with the observer looking 120° in azimuth away from the lamp, which is out of the picture at the right. Focal length = 15 mm.

one good one and may have some insights, and perhaps we will all see more of them, now being duly sensitized. In any event, the above discussion is probably not so relevant to the parhelic circle for small lamp elevations like the $\Sigma = 10°$ of the previous diagrams. For small Σ the parhelic circle surface is nearly planar, and, moreover, the rubber band for small Σ has too large an angular diameter to be seen as a whole by the observer. The streetlight parhelic circle should be more interesting for larger Σ.

It should be relatively easy to identify the common halos when they are formed in streetlights, if they are eye-connected. Each eye-connected halo curve should appear as a curve in space that ends at a point that is its classical instance, and each eye-connected halo surface should appear as a surface in space that ends at a curve that is its classical instance. Streetlight halos that are not eye-connected, on the other hand, do not give away their identities so easily. Compare, for example, the subparhelion caustic and the superparhelion caustic in Fig. 1.3, the former being eye-connected, the latter not.

The eye-connected halos for a given lamp elevation Σ are exactly those halos whose classical instances exist (are non-empty) for sun elevation Σ. The superhalos in Fig. 1.5, with $\Sigma = 10°$, are not eye-connected, nor would they be eye-connected for other positive Σ. So they do not exist classically unless the sun is below the horizon. That's why you don't find, say, superparhelia in the halo literature.

1.3 Lamp below the observer

All of the above diagrams had lamp elevation $\Sigma = 10°$. It is fascinating to let Σ vary and see how the diagrams change, and I will do so in Chapter 3, but let's consider here just one new lamp elevation, namely, $\Sigma = -10°$. So the lamp is now lower than the observer.

For $\Sigma = -10°$ all you do[3] is turn upside down all the diagrams for $\Sigma = +10°$. Abstractly, then, the two cases are equivalent. In practice, however, there is a difference, and this difference explains why I had never seen anything like the seemingly strange halo in Jukka's photo. The difference is that the observer is usually standing on the ground, with no crystals below him, so that the sub-halos cannot be activated. In plainer language,

[3]However, although I have not done so here, one can argue that the halo notation should change when $\Sigma < 0$: The symbols C_1 and C_2 should perhaps be transposed, so that C_1 becomes the subparhelic circle and C_2 the superparhelic. That would keep C_1 always on the lamp side of space, and it would keep C_2 on the observer side, regardless of whether Σ were positive or negative. Similarly for P_1 and P_2, etc.

you usually don't see sub-halos. The superhalos in Jukka's photo, Fig. 1.1, would not have looked so exotic to me, had I ever seen the subhalos predicted by Fig. 1.4. By pointing his spotlight upwards, making Σ negative, Jukka made superhalos that were inverted versions of the hard-to-see subhalos for Σ positive. Although I made no effort to match the parameters of Jukka's photo when I made Fig. 1.10 and Fig. 1.11, it is pretty clear that the halo that puzzled me was, as many of you thought, the superparhelion, but with negative lamp elevation. (I think your terminology was different. My naming convention is that the halos on the upper sheet are always the superhalos, and those on the lower sheet are always the subhalos, regardless of whether the lamp is above or below the observer.)

The observers remarked that the display seemed stunningly three-dimensional, more so than in the average streetlight display. I think I now see why, by comparing Fig. 1.5 and Fig. 1.10. The superparhelion surfaces are curved the same in both figures. Usually, however, with $\Sigma > 0$, as in Fig. 1.5, the superparhelia are far from the observer (and close to the lamp), whereas with $\Sigma < 0$, as in Fig. 1.10, the superparhelia are close to the observer. That is perhaps what made the curvature more perceptible.

The halo diagrams in this chapter are obviously hugely idealized, and they would need many modifications to make them realistic. The diagrams assume zero tilts for the crystals.[4] They ignore the crucial and difficult question of how the spatial nature of streetlight displays is perceived by the brain. And the diagrams contain hardly any intensity information. For example, they depend only on the angle of elevation of the lamp, not on the observer's distance from the lamp, but the distance matters in practice, since, among other things, you run out of crystals as you get close to the lamp.

[4]The fact that the superlamp in Fig. 1.1 extends below its classical position—and in fact all the way to the lamp—is a sign that the crystal orientations were not perfect.

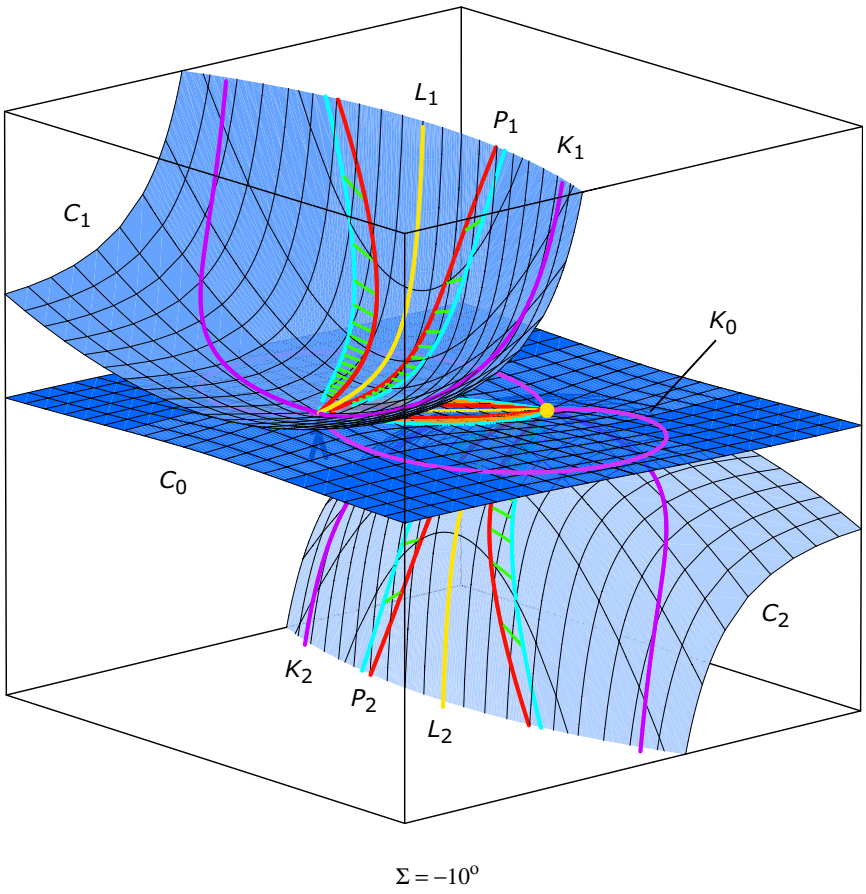

Figure 1.10: Same as Fig. 1.5 but with $\Sigma = -10°$ instead of $\Sigma = +10°$, so that the lamp is below the level of the observer.

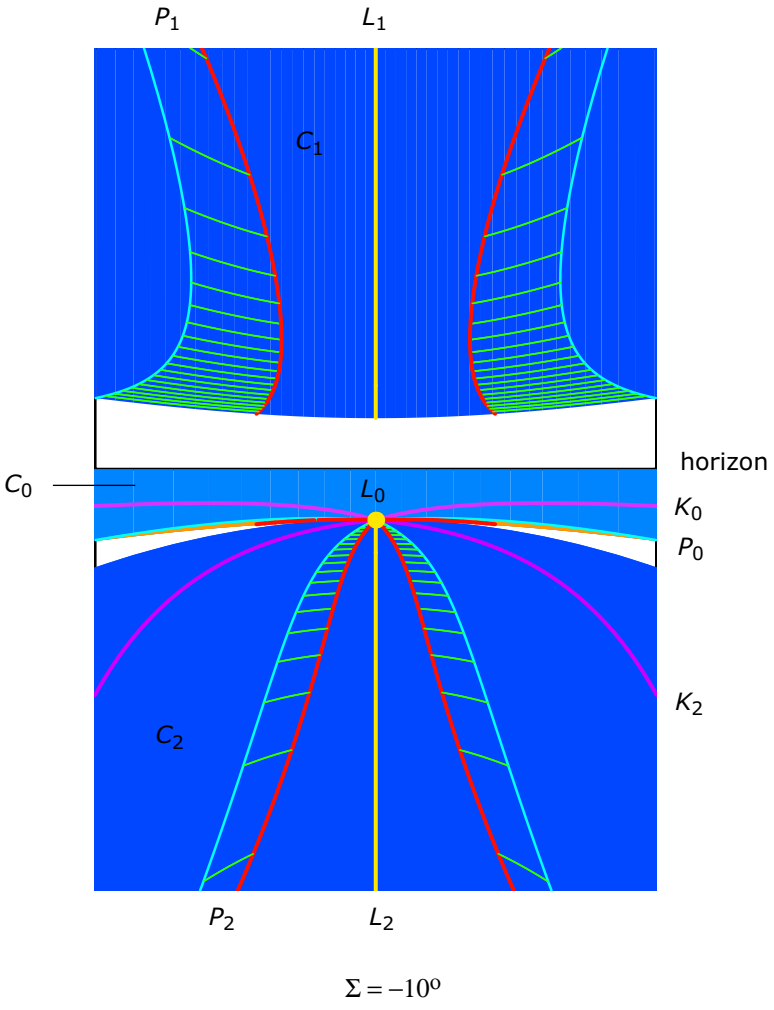

Figure 1.11: The halos of Fig. 1.10 ($\Sigma = -10°$) as if photographed by the observer with a 12 mm rectilinear lens.

Chapter 2

Superlamp and Sublamp

The simplest and most common streetlight halos are the sublamp and superlamp. Their classical instances are the subsun, which is common, and the supersun, which nobody has ever heard of, for reasons that will become apparent. Although the sublamp and superlamp are simple, they illustrate many of the features seen in more complicated spatial halos.

The sublamp and superlamp are both shown in Fig. 2.1, where they are depicted as geometrical curves in space. The same curves, as they would appear to the observer, are shown in Fig. 2.2.

2.1 Meaning of the spatial diagrams

The halo curves in Fig. 2.1 are not real physical objects. The observer cannot leave his (or her) position, wander around, and view the curves from other perspectives, say that of the reader. But the curves do show how the observer might report the appearance of the halos.[1] Precisely, they show the regions of space where crystals can send halo light to the observer's eye. The crystals that make the super- and sublamps are horizontal plates, and the ray paths are reflections off the horizontal faces of the crystals, as in Fig. 2.3. Points on the superlamp and sublamp curves are exactly the places where horizontal crystals can send light to the eye by means of reflections from their horizontal faces.

To understand this chapter you do not need to know a formula for the superlamp and sublamp curves, but I am going to do the calculation anyway, because it makes more concrete the meaning of diagrams like Fig. 2.1. In

[1] I am much less certain about this remark for the super- and sublamp than for other spatial halos. More later.

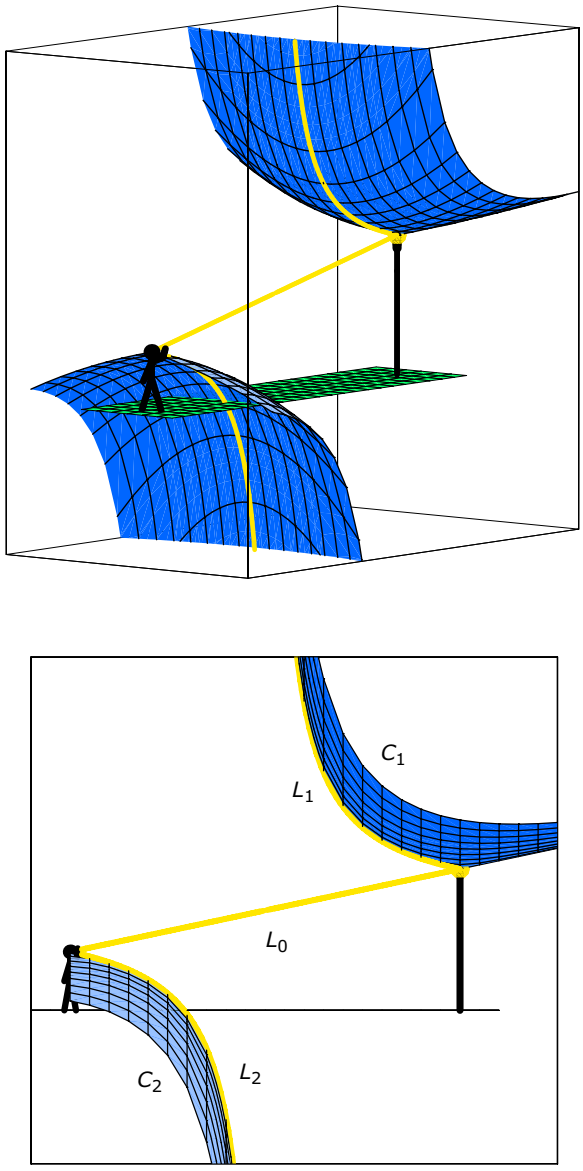

Figure 2.1: A simple streetlight halo display. The yellow curves are the superlamp L_1, the sublamp L_2, and the 'lamp halo' L_0. The blue surfaces are the superparhelic circle C_1 and the subparhelic circle C_2. For completeness the subparhelic circle and the sublamp are shown extending below the ground. The elevation angle of the lamp with respect to the observer's eye is $\Sigma = 12°$.

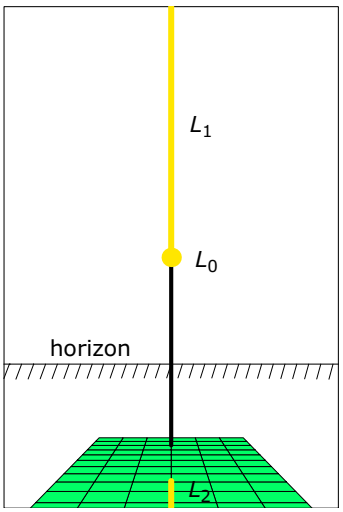

Figure 2.2: The lamp, superlamp, and sublamp of Fig. 2.1 as if photographed by the observer with a 35 mm rectilinear lens. The camera cannot sense the spatial nature of the display.

Fig. 2.4, the point (y, z), $-1 < y < 1$, is on the superlamp or sublamp curve exactly when $\sigma^* = \sigma$, as in the figure. Equivalently,

$$\tan \sigma^* = \tan \sigma \tag{2.1}$$

$$\frac{z + h/2}{y + 1} = \frac{z - h/2}{1 - y} \tag{2.2}$$

$$yz = \frac{h}{2} \tag{2.3}$$

Equation (2.3) describes a hyperbola that passes through the lamp $(1, h/2)$ and the eye $(-1, -h/2)$ and that has the y-axis and z-axis as asymptotes. Adding the fine print and putting everything in three dimensions, you find that the superlamp and sublamp curves are given by

$$yz = \frac{h}{2}, \quad x = 0, \quad -1 < y < 1 \tag{2.4}$$

The calculation reveals some of the limitations of diagrams like Fig. 2.1. No considerations of intensity went into the calculation. The crystal orientations were assumed to be perfect, with zero tilts. We did not worry about questions of perception. And so forth.

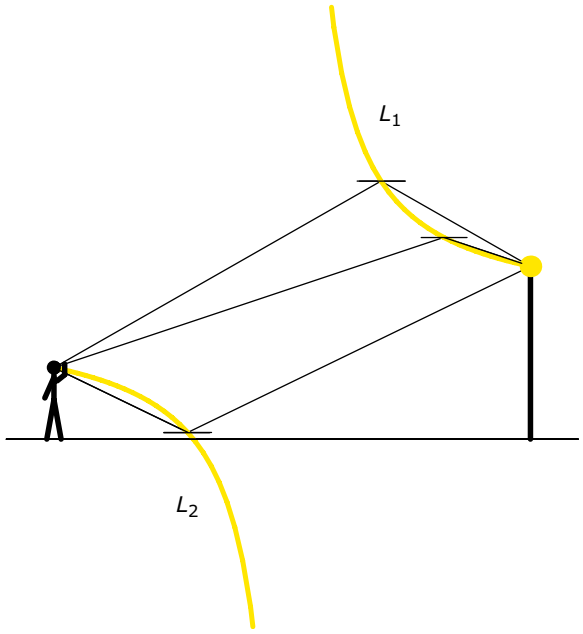

Figure 2.3: Illustrating the meaning of the superlamp and sublamp curves L_1 and L_2 (yellow). They consist of the locations where little horizontal mirrors—or horizontal plate crystals—can reflect light to the observer's eye.

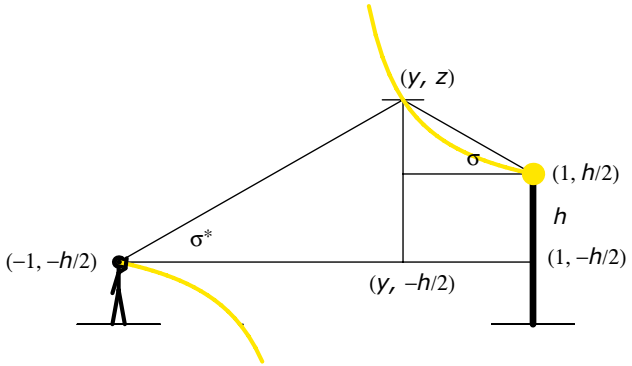

Figure 2.4: Diagram for the derivation of Eq. 2.3, which characterizes the superlamp and sublamp curves (yellow). The point (y, z) is on one of the curves exactly when $\sigma^* = \sigma$. The plane of the paper is the yz-plane, and the origin of coordinates (not shown) would be at the midpoint of a line segment joining the lamp and the observer.

But, again, you do not need to know Eq. (2.4); the main point is just that you can calculate the regions in space where crystals can send light to the eye in specified ways. That is what Fig. 2.1 is showing in the case of the superlamp and sublamp. Thus the yellow curves are what the observer will perceive as being the halos. But the reader, from his own perspective, will never see them.

Figure 2.1 also shows the 'lamp halo.' That is the region in space where crystals can send light to the eye undeviated. Thus the lamp halo is just the line segment between the eye and the lamp. To the observer it appears as the lamp itself. It is not an interesting halo.

2.2 Classical superlamp spot

We can use the superlamp as an example to review the properties of eye-connected halos.

Recall from Section 1.2 that a halo curve is eye-connected if the curve reaches the observer's eye. With the lamp above the observer, as in the left-hand diagrams of Fig. 2.5, the superlamp curve is not eye-connected and there are no surprises. With the lamp below the horizon, however, as in the right-hand diagrams, the superlamp becomes eye-connected. As a result, the observer can look tangentially along the curve as it leaves the eye, thus seeing sparkles from many crystals, all more or less in a line. Under favorable conditions the observer will perceive a spot of some sort in that direction. His brain, in some mysterious way, will decide how far away to locate the spot in space.

The superlamp in the right-hand diagrams of Fig. 2.5 is typical of eye-connected halo curves in general. To the observer they appear to end at some point in space. At that point the observer often sees a relatively bright spot, not necessarily well-defined. Crystals making the spot are close to the eye, and so individual sparkles are apt to be seen in the spot, and indeed the spot sometimes appears only as a concentration of sparkles.

The direction of the spot is not random. The crystals that make it—the ones closest to the eye—are exactly the crystals that make the classical instance of the halo curve. (Imagine the lamp in the diagram to be the sun and the observer to be in his usual place on Earth, seeing halos in sunlight; the halo-making crystals would be in a tiny space around the observer.) Thus in Fig. 2.5 the yellow spot labelled 'L_1 at the eye,' where the superlamp curve appears to end, must be in the direction of the supersun—the classical instance of the superlamp. With $\Sigma = -5°$, as in this example, the supersun

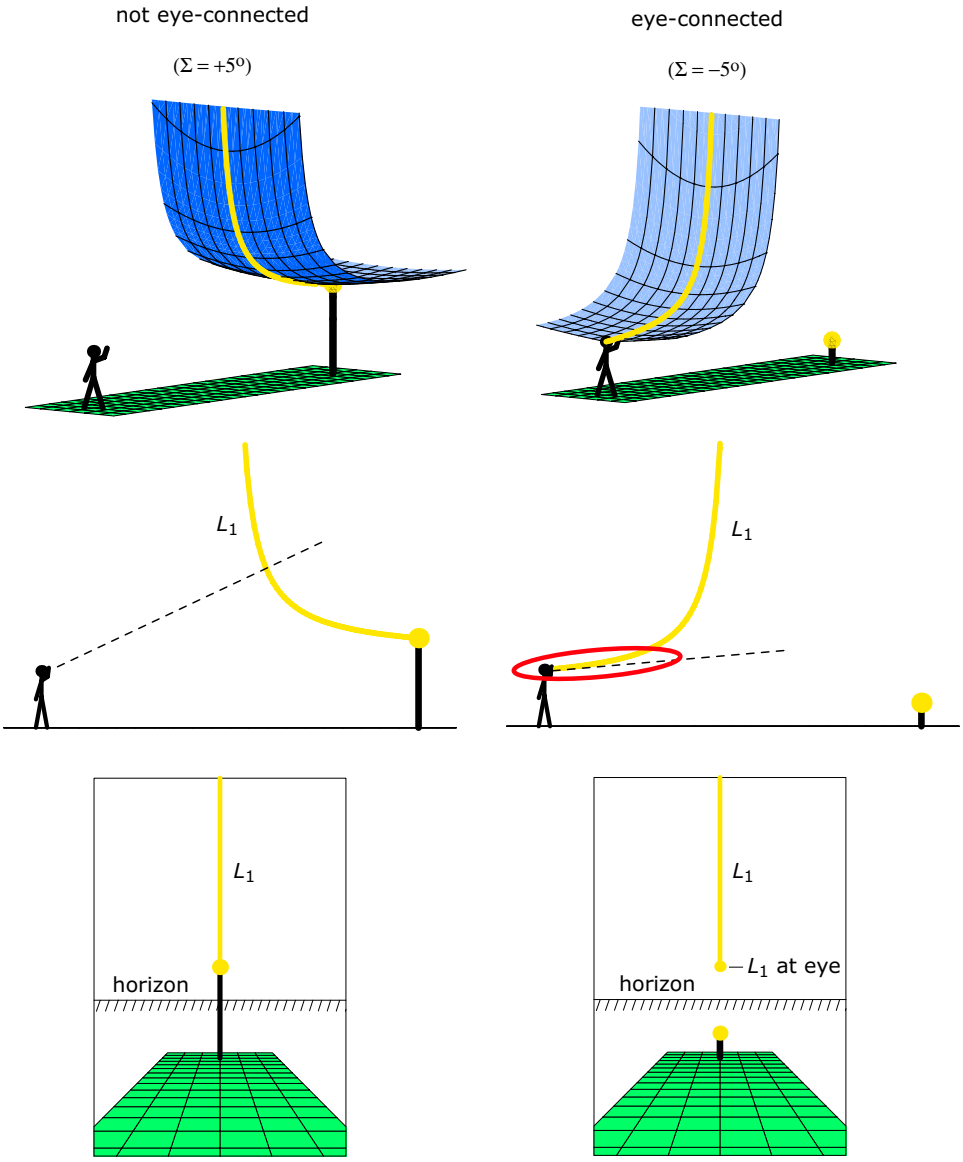

Figure 2.5: Eye-connected versus not eye-connected. With lamp elevation $\Sigma > 0$ (left), the superlamp curve L_1 is not eye-connected. The observer looks at the curve transversely and sees relatively few crystals. With $\Sigma < 0$ (right) the superlamp is eye-connected. The observer is able to look along the curve tangentially (red) as it leaves the eye, and in that direction he sees many crystals. This is the direction of the classical superlamp point, or 'L_1 at the eye,' where, to the observer, the superlamp curve appears to end.

would be a white spot five degrees above the horizon.

Hardly anyone has heard of a supersun. That's because the sun—or other parallel light source[2]—needs to be below the horizon in order for the supersun to be non-empty.[3] So it sounds odd to say 'classical superlamp,' but I am going to do so anyway. The spot that we have been talking about I will call the classical superlamp (point or spot), since, after all, that is what it is. Originally I called it simply the superlamp spot, but we need some terminology that clearly distinguishes it from the zenith spot (Section 2.3).

In general, a halo curve is eye-connected at lamp elevation Σ if and only if its classical instance is non-empty for sun elevation Σ. So to say that the superlamp curve is not eye-connected for positive lamp elevations is the same as saying that the supersun is empty for positive sun elevations.

2.3 Zenith spot

With lamp elevation Σ negative, the superlamp curve is, as I said, eye-connected. If the observer looks in the direction of the curve where it leaves the eye, he is looking tangentially along the curve and can hope to see the classical superlamp spot.

There is another portion of the curve that the observer can view nearly tangentially, and this is true for positive as well as negative Σ. That portion is the upper part of the curve. The superlamp curve, recall, is a hyperbola with a vertical asymptote. All vertical lines, as seen by the observer, converge on the zenith, and hence so does the superlamp curve. In the direction of the zenith the observer is looking along the superlamp curve nearly tangentially and thus will sometimes see a relatively bright spot. Some of you folks have called it the zenith spot, and I will do the same. It will appear to be an endpoint of the superlamp curve, but it is not the same as the endpoint that is the classical superlamp spot. Thus the superlamp can masquerade as an eye-connected curve even when it is not. But the disguise is not perfect; it is revealed by the location of the spot.

[2]Jenni Elina Holopainen photographed a supersun formed in sunlight reflected from a lake surface. See the Halo Reports website [4] for Thursday, May 18, 2006.

[3]This assumes perfectly horizontal crystals. If you allow less than perfect crystal orientations, then the supersun becomes a common item, since a pillar above the sun is a poorly formed supersun.

2.4 Scale matters

Scale is irrelevant to the halo curves themselves. If you start with, say, Fig. 2.1, and you double the distance from the observer to the lamp, keeping Σ the same, then to make a new diagram you just double all the distances in the figure. In practice, however, the scale has a big impact on what the observer actually sees. The zenith spot is one illustration.

The left-hand diagram in Fig. 2.6 is the same as the left-hand diagram in Fig. 2.5. Lamp light reaching the upper part of the superlamp curve, where the zenith spot originates, is apt to be too weak for the zenith spot to materialize. In the right-hand diagram in Fig. 2.6, however, with the observer much closer to the lamp, the upper part of the hyperbola will be well-lit and the observer will see the zenith spot.

But it seems you can't get something for nothing. By standing right next to the lamp, as in the right-hand diagram, the observer probably loses his chance to see the lower part of the superlamp, the part that appears to be adjacent to the lamp. He just runs out of crystals in the relevant section of the curve.

Here is another way in which scale matters: I said that an eye-connected curve will often appear to end at a point in space that is accentuated by sparkles. That's fine if the observer is reasonably well lit by lamp light, so that the crystals near him—the ones creating the sparkles—are also well lit. But if the observer backs away, out of the main glare of the lamp, the sparkles from nearby crystals will be too weak to be seen. The halos need not disappear entirely, but they will be less adorned by sparkles.

2.5 Observing the superlamp

I have not had much chance this winter to try out the ideas of this chapter on real streetlight halo displays. I have no idea to what extent they will have to be modified by the harsh realities of intensity factors, imperfect crystal orientations, matters of perception, and the like. You folks who can hang around the snow guns at ski areas and order displays to your liking will have more opportunities to test these ideas. A decent display—with small crystal tilts—will be needed for a meaningful test.

And the tests will be trickier without an omnidirectional lamp, which is hard to find. Using a spotlight, with its narrow beam, you will need to think ahead about what portion of the superlamp curve you are trying to illuminate.

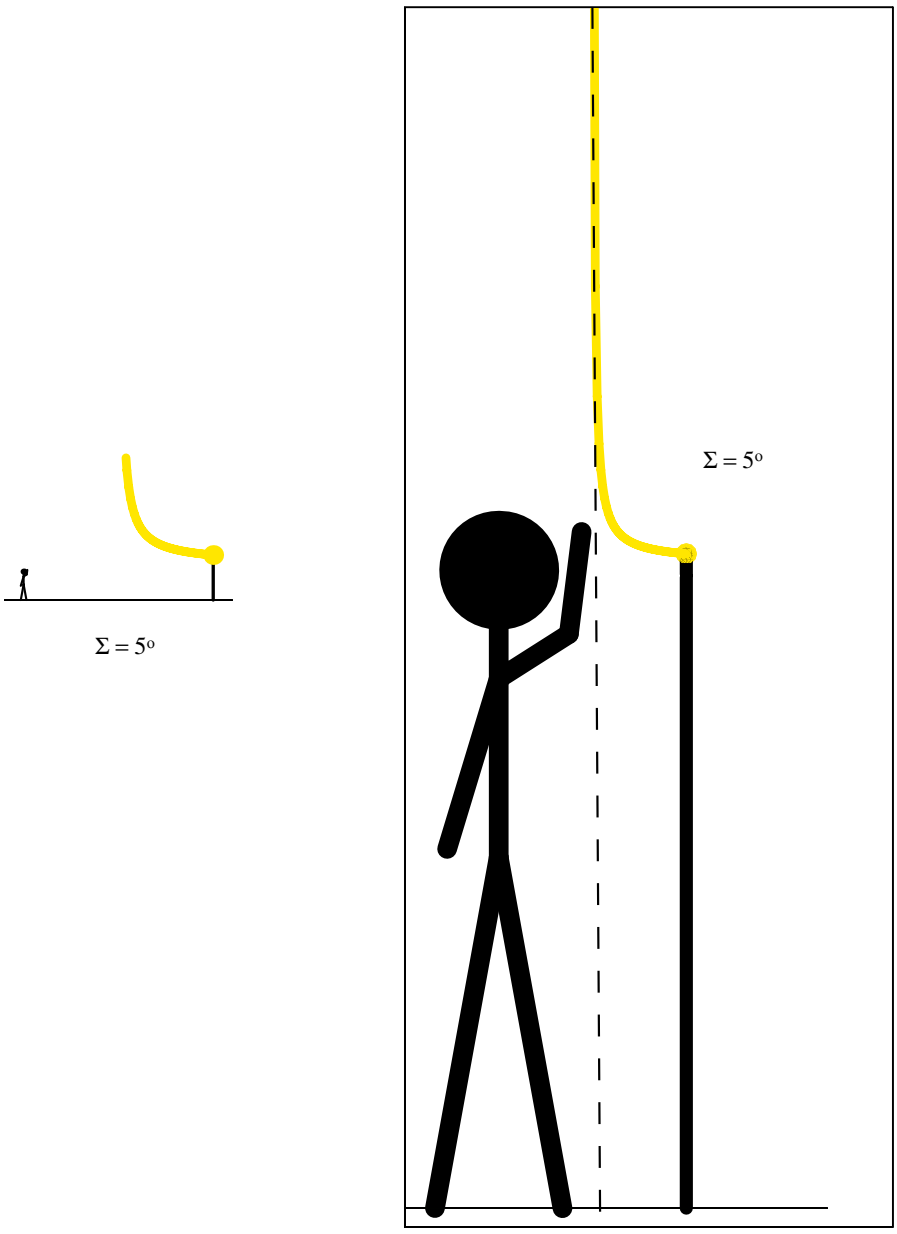

Figure 2.6: Illustrating that scale matters. The lamp elevation is $\Sigma = 5°$ in both diagrams, so the geometry of the superlamp curve is the same in both. But the observer in the right-hand diagram has a better chance of seeing the zenith spot, since in that diagram the upper portions of the superlamp curve should be better lit. Looking nearly tangentially along that part of the curve, the observer should be able to see the zenith spot—a bright spot very near the zenith.

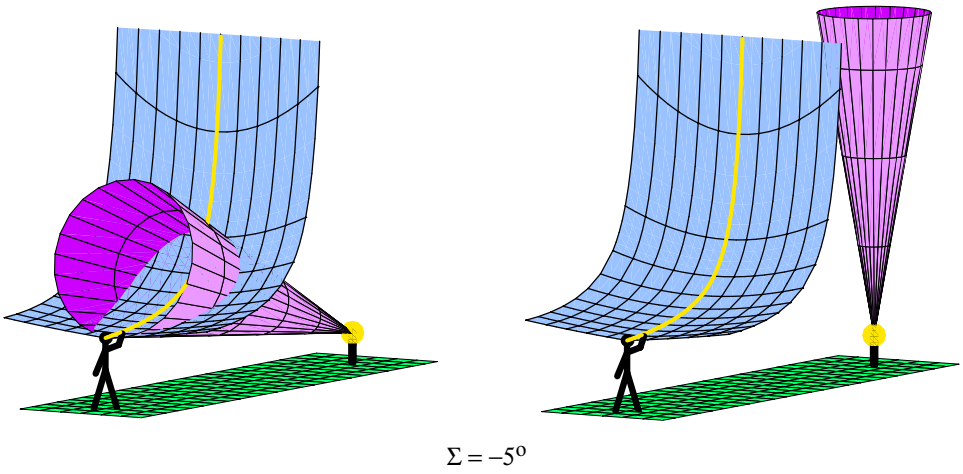

Figure 2.7: Observing the classical superlamp spot with a spotlight. The conical beam from the spotlight needs to fall on the portion of the curve near the eye, as at the left. The observer in the right hand diagram will not see the classical superlamp spot. ($\Sigma = -5°$)

Recall that there are two spots on the celestial sphere that are associated with the superlamp curve, namely, the classical superlamp spot and the zenith spot. The classical superlamp spot can only occur when Σ is negative. In that case the spot—as seen in a photo—would be directly above the lamp at elevation $|\Sigma|$, and would have a tail extending upward. The zenith spot would be near the zenith, with the tail extending downward. The classical superlamp spot, the tail, and the zenith spot together make up the superlamp as seen by the observer. It will be fun to see if you can get all three to appear simultaneously. My sense is that you will need to place the lamp low but not too low. That is, you want the lamp low enough so that the classical superlamp spot is high, but not so low that the classical superlamp spot merges with the zenith spot. With $\Sigma = -80°$, for example, the classical superlamp spot would be ten degrees from the zenith, and the superlamp would extend from the zenith spot down towards the classical superlamp spot. But you will have to experiment not only with the lamp elevation but with the observer-lamp distance, as explained in connection with Fig. 2.6.

Another suggestion, easy to make from an armchair, less easy at night in the cold: For halo chasers in a group, try to generate zenith spots from

two or three spotlights simultaneously. The spots should be close, even coincident, but each tail should point to its own spotlight.

For many years I have known that the superlamp and sublamp halos constituted a hyperbola in space, and I was always puzzled that I could not detect the curvature in reality. (Remember that the observer does not have the benefit of the reader's perspective, as in, say, Fig. 2.1; for the observer there is no side view.) The notion of eye-connected had escaped me, and I did not understand the distinction between positive and negative lamp elevation (Fig. 2.5). I see now that for positive lamp elevations Σ—the normal streetlight scenario—the superlamp curve is not eye connected. But with Σ negative, the superlamp curve should take on the characteristic look of an eye-connected curve. It should start at a spot in space and then trail away upward, falling away from the observer. If the observer is located at a suitable distance from the lamp, then sparkles along the curve should accentuate the falling away, with the sparkles diminishing at distance. And, unlike the superlamp with $\Sigma > 0$, all this will be happening with the curve nearer to the observer than to the lamp. But you will have to experiment with the lamp elevation. It needs to be negative, but you don't want it too negative, since then there is very little curvature in the curve, even in the side view. I am guessing that moderately negative lamp elevations will be most favorable and most interesting. An omnidirectional light would help, since then you do not have to worry about the spotlight being aimed correctly, at the crucial part of the curve. It may turn out, however, that the curvature is still not detectable, since the curve is bending directly toward or away from you. I am eager to look again next winter.

Examples

Figure 2.8 shows the zenith spot with tail extending down toward the lamp, which is out of the photo but whose location is revealed by the strong beam emanating from it. The zenith spot and the tail together constitute the superlamp here. By the way, the beam is a different sort of creature than the halos. It arises in badly formed and badly oriented crystals, or in dust particles, etc., all of which send out light in more or less random directions, so that the beam is more akin to a real physical object than to the halos. You can walk around the beam and watch it, and it will appear to stay put, as if it were, say, a tree. Not so with the halos.

Figure 2.9 shows the superlamp with a strong classical superlamp spot. Of course there is much else to marvel at in the photo.

You might think that it is a major undertaking to find examples of eye-

Figure 2.8: Superlamp, including the zenith spot. Pälkäne, Finland, January 16-17, 2009. ©Jari Luomanen.

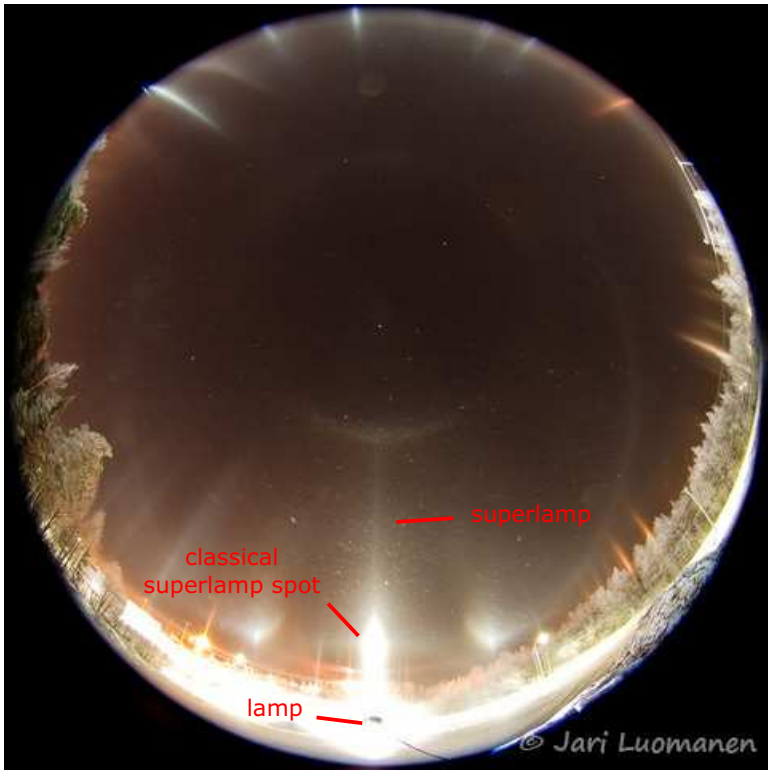

Figure 2.9: Superlamp, including the classical superlamp spot. The lamp is slightly lower than the camera. Tampere, Finland, January 2-3, 2009. ©Jari Luomanen.

connected and non-eye-connected curves in reality. But the most mundane plate crystal display in a streetlight is apt to contain one example of each. Standing on the sidewalk (with $\Sigma > 0$), you perceive a streak of light extending straight up from the lamp. That is the superlamp. Then looking somewhere below the lamp, you see a concentration of sparkles between yourself and the sidewalk. That is the sublamp. The sublamp is eye-connected, the superlamp is not. See how different they appear.

2.6 H at the eye, H at infinity

For the superlamp curve when $\Sigma < 0$ there can be both the classical superlamp spot and the zenith spot. The classical superlamp spot I will sometimes refer to as the 'superlamp at the eye,' and the zenith spot I will sometimes refer to as the 'superlamp at infinity.' To the casual observer the two spots can look similar, but they arise in quite different ways, as we have seen. The existence of the superlamp at the eye depends on the superlamp curve being eye-connected, as it is for $\Sigma < 0$. The existence of the superlamp at infinity depends on the superlamp curve being unbounded, meaning that you cannot fit the curve inside any one sphere, no matter how big you make the sphere.

The superlamp suggests how other halos that are eye-connected or unbounded should be treated. If H is any eye-connected halo curve, then define 'H-at-the-eye' to be to be the point on the celestial sphere in the direction that the curve H leaves the eye. And if H is an unbounded halo curve, then define 'H-at-infinity' to be the point on the celestial sphere in the direction that H 'arrives at infinity.'[4] So H-at-infinity is the conventional vanishing point of the curve H; it is where you look to see the curve receding into the distance. And H-at-the-eye is the point that is the classical instance of the halo H. But I cannot think of any familiar everyday analog for it.

If S is a halo surface instead of a curve, then define 'S at the eye' to be the set of all points H-at-the-eye for which H is an eye-connected curve in S. Similarly, define 'S at infinity' to be the set of all points H-at-infinity for which H is an unbounded curve in S. So S at the eye is the curve that

[4] Better:

$$H\text{-at-the-eye} = \lim_{|\mathbf{X}-\mathbf{E}|\to\mathbf{0}} \frac{\mathbf{X}-\mathbf{E}}{|\mathbf{X}-\mathbf{E}|}$$

$$H\text{-at-infinity} = \lim_{|\mathbf{X}-\mathbf{E}|\to\infty} \frac{\mathbf{X}-\mathbf{E}}{|\mathbf{X}-\mathbf{E}|},$$

where $\mathbf{E}$ is the position of the observer's eye and $\mathbf{X}$ is a generic point on H.

is the classical instance of the halo S, and S at infinity is the curve where the surface S appears to recede into the distance,

Figure 1.4 has some examples of surface halos at infinity and at the eye, though they are not labelled there. In that figure the upper boundary of the parhelic circle C_0 is C_0 at the eye, the classical parhelic circle. The lower boundary is C_0 at infinity, the horizon. The boundary of the subparhelic circle C_2 that is visible in the figure is C_2 at the eye, the classical subparhelic circle, and the boundary of the superparhelic circle C_1 is C_1 at infinity.

Figure 2.10 is more exotic, with the lamp well below the level of the observer. The superlamp curve L_1 and the right and left 120° superparhelion curves K_1 are eye-connected and unbounded, and so each appears to the observer to have two endpoints, neither of which is the lamp. These points are L_1 at the eye, L_1 at infinity, K_1 at the eye, and K_1 at infinity. The superparhelic circle surface C_1 is also eye-connected and unbounded, and so it appears to have two boundary curves, namely, C_1 at the eye and C_1 at infinity. But I suspect that getting all of these features to show up in a real display would be a very tall order. I wonder if even the two boundary curves of the superparhelic circle (surface) could be made to show up at once. It would be a lot tougher than just getting the two endpoints of the superlamp (curve) to materialize.

By the way, the only direction in which the observer can look exactly tangent to the parhelic, superparhelic, or subparhelic circle surfaces is precisely in the direction of the lamp, which is not an interesting case. So the only place where the geometry of these surfaces might lead to high perceived intensity is at the eye and at infinity; there is no analog of 'at-the-eye' or 'at-infinity' on the surfaces themselves. By way of contrast: when the observer looks at the 'vee' of the helic arc (surface) in, say, Fig. 10.10, he is looking tangentially along the surface, neither at the eye nor at infinity.

One might think that the behavior at infinity would be irrelevant, but in practice 'infinity' need only be far enough away so that the direction of the line from observer to the halo curve (or surface) is not changing much as the observer looks farther and farther out on the curve. 'Infinity' therefore depends on the distance from the observer to the lamp. Scale matters.

2.7 Caveat: The Lamp is Not a Point

This section was written long after much of the rest of this tome, because I did not get the point soon enough. But it bears on the appearance of all classical halos in streetlights.

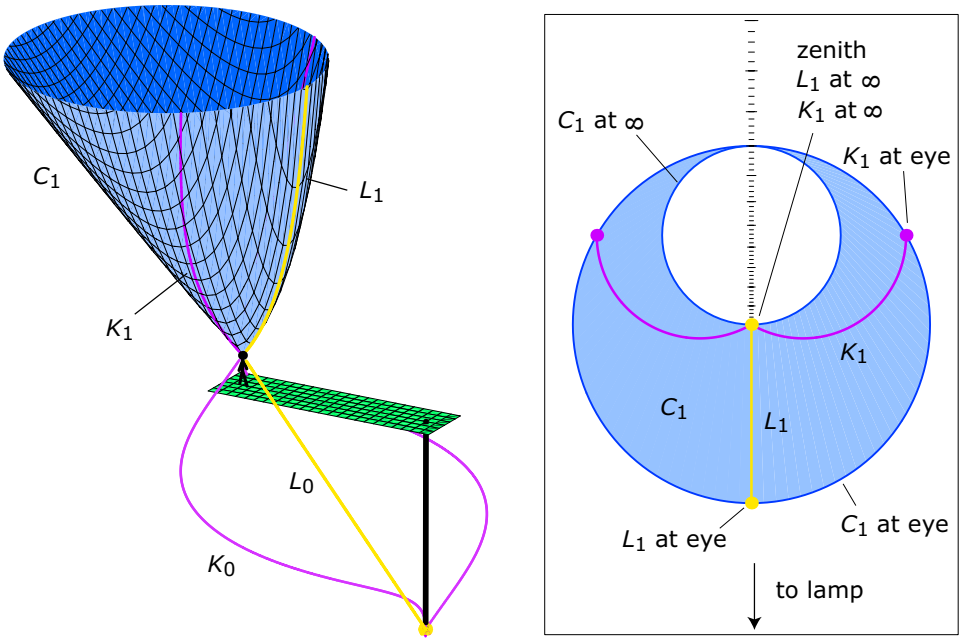

Figure 2.10: (*Left*) Streetlight halo display with $\Sigma = -50°$. The superlamp L_1, the 120° superparhelia K_1, and the superparhelic circle C_1 are eye-connected and unbounded, so that they each have instances at both the eye and at infinity. The lamp line L_0 and the 120° parhelia K_0 are also shown, but the parhelic circle is omitted, as is the subparhelic circle and its halos. (*Right*) The same as if photographed by the observer with a 12 mm rectilinear lens aimed straight up.

As an example, let's consider the right parhelion caustic curve. For low to moderate lamp elevations the curve is eye-connected and therefore has a classical instance. If the lamp (bulb) is a point, and if—as we normally assume—the halo-making crystals are perfectly horizontal, then the parhelion curve is a true mathematical curve. The linear density of crystals along the curve ought to be constant, and the classical instance of the parhelion—a point—ought to be relatively bright, since the observer looks tangentially along the curve to see it.

But the lamp is not a point—it has some positive radius. So there is not just one parhelion caustic curve but rather a whole bunch of them, each starting from a different point on the lamp and then ending at the eye. Together they make up the parhelion caustic. The caustic is no longer a curve but rather a solid tube, thick near the lamp and then tapering down to a point[5] at the eye. See Fig. 2.11.

Due to the tapering, there will be a relative dearth of halo-making crystals near the eye, precisely in the region where the classical instance of the halo is trying to form. It seems to me that there will always be a tension between the tapering, on the one hand, and the extent of the straightness of the halo near the eye, on the other. The straightness encourages the classical instance of the halo, the tapering inhibits it.

Figure 2.12 is like Fig. 2.11 but shows the superlamp and sublamp instead of the parhelion. The idea is the same; the sublamp tube peters out at the eye, and this weakens the classical sublamp.

There are probably many factors that combine to determine whether the classical instance of a halo will materialize. I am at a loss to predict under what circumstances it will do so.

Acknowledgement

I want to thank Jari Luomanen for some very helpful email exchanges, especially regarding the interpretation and the circumstances of his photo in Fig. 2.8.

[5]The eye, like the lamp, is not quite a point either, but the eye is apt to be small compared with the lamp. Perhaps we can indeed treat the eye as a point.

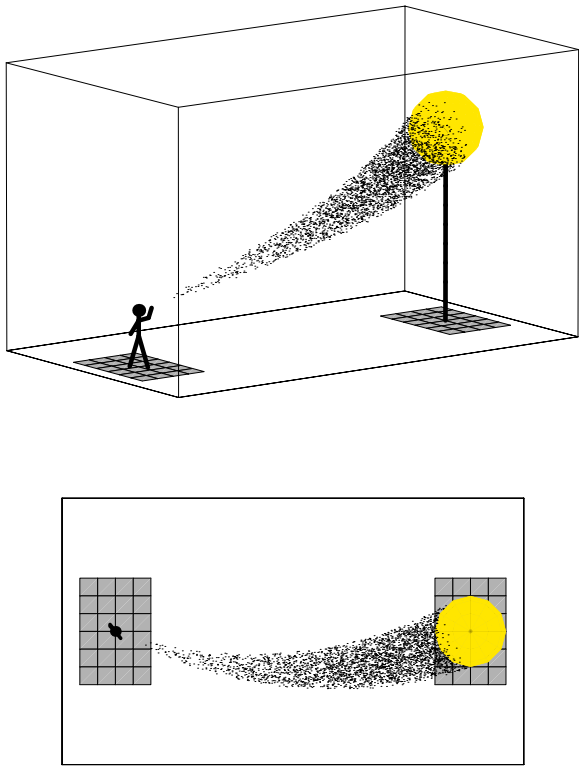

Figure 2.11: Right parhelion caustic formed in a lamp with non-zero bulb radius. Had the bulb been a point (zero radius), the parhelion caustic would have been a curve. Instead it is a solid curved tube starting at the lamp and tapering down to a point at the observer's eye. The tapering should weaken the classical parhelion caustic, since it means that there are relatively few halo-making crystals near the eye.

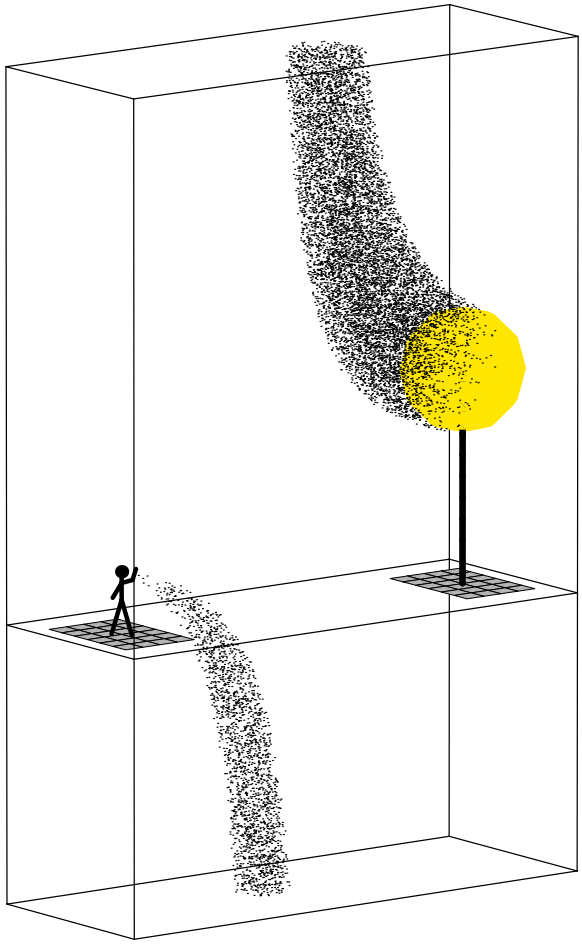

Figure 2.12: Superlamp and sublamp formed in a lamp with non-zero bulb radius. Like the parhelion caustic in Fig. 2.11, the sublamp tapers down to the eye. This weakens the classical sublamp.

Chapter 3

Mathematics of the Parhelia

This chapter describes how to calculate the streetlight parhelia, superparhelia, and subparhelia, and hence how to make spatial halo diagrams like those in Chapter 1. It does make some mathematical demands, and not everyone will want to plow through the mathematics, but the figures here might be worth a look anyway. They give key insights into spatial halos, and they show the dependence of the parhelia, superparhelia, and subparhelia on lamp elevation. This dependence is interesting in its own right, and it is fundamental for understanding other refraction halos.

This chapter is not about intensities or about what the halos look like in reality. It is just about how you calculate the halos, that is, how you calculate the surfaces in space where the halo-making crystals refract light to the observer's eye.

3.1 Combined parhelic circle

I will assume that the lamp height h is positive. There is no loss of generality in doing so, since the case $h < 0$ follows from $h > 0$ by interchanging the observer and the lamp.

The *combined parhelia* $\cup P_i = P_0 \cup P_1 \cup P_2$ consists of the parhelia, superparhelia, and subparhelia together, and the *combined parhelic circle* $\cup C_i = C_0 \cup C_1 \cup C_2$ consists of the parhelic, superparhelic, and subparhelic circles together. I don't much like these new terms, but we need something of the sort.

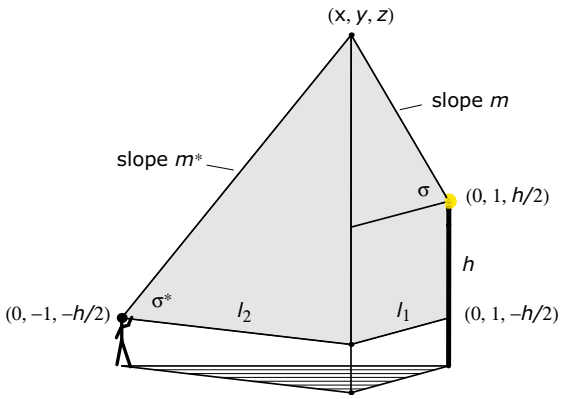

Figure 3.1: Diagram for the derivation of the equations for the parhelic circle, subparhelic circle, and superparhelic circle [Eq. (3.1)]. The point (x, y, z) is on the subparhelic circle or superparhelic circle when, as here, $\sigma^* = \sigma$, and it is on the parhelic circle when $\sigma^* = -\sigma$. The yz-plane (not shown) is vertical and passes through the observer and the lamp, with the origin of coordinates midway between the eye and the lamp. The ground surface (stippled) is logically irrelevant.

Since $\cup P_i$ is a subset of $\cup C_i$, I will treat $\cup C_i$ first. In Fig. 3.1:

$(0, \quad 1, \quad h/2)$	position of lamp
$(0, -1, -h/2)$	position of observer's eye
Σ (not shown)	elevation angle of lamp from observer's eye
$h = 2 \tan \Sigma$	height of lamp measured from level of observer's eye
$l_1 = \sqrt{x^2 + (y-1)^2}$	horizontal distance from (x, y, z) to lamp
$l_2 = \sqrt{x^2 + (y+1)^2}$	horizontal distance from (x, y, z) to observer
σ	elevation angle of (x, y, z) as seen from lamp
σ^*	elevation angle of (x, y, z) as seen by observer
$m \ = \tan \sigma$	slope of segment from lamp to (x, y, z)
$m^* = \tan \sigma^*$	slope of segment from eye to (x, y, z)

So the elevation angle σ and slope m are positive if (x, y, z) is above the lamp, negative if below. And σ^* amd m^* are positive if (x, y, z) is above the eye, negative if below.

The point (x, y, z) is on the combined parhelic circle $\cup C_i$ when $|\sigma^*| = |\sigma|$

or, equivalently,

$$|m^*| = |m| \quad \text{(equation for } \cup C_i\text{)}. \tag{3.1}$$

Specifically,

$$m^* = -m \quad \text{for } C_0 \tag{3.2}$$

$$m^* = m \text{ and } \begin{cases} m^* \geq 0 & \text{for } C_1 \\ m^* \leq 0 & \text{for } C_2 \end{cases} \tag{3.3}$$

Since

$$m = \tan\sigma = \frac{z - h/2}{l_1} \tag{3.4}$$

$$m^* = \tan\sigma^* = \frac{z + h/2}{l_2}, \tag{3.5}$$

then

$$z = \frac{h}{2}\frac{l_2 - l_1}{l_1 + l_2} \quad \text{for } C_0 \tag{3.6}$$

$$z = \frac{h}{2}\frac{l_1 + l_2}{l_2 - l_1} \quad \text{for } C_1 \cup C_2 \quad (h \neq 0). \tag{3.7}$$

So if we can understand the combined parhelia as seen from above, that is, as projected onto the xy-plane, then we can use Eqs. (3.6) and (3.7) to construct them on the combined parhelic circle $\cup C_i$. That is the plan. But to understand the projected views requires the Bravais refractive index (Section 3.2.2), which depends on the slope m, so we need to think about how m varies on $\cup C_i$.

We can express

$$\cup C_i = \bigcup_{k \geq 0} C^k(h), \tag{3.8}$$

where $C^k(h)$ is the set of points (x, y, z) such that

$$|m| = k \tag{3.9}$$

$$|m^*| = k \tag{3.10}$$

Thus $C^k(h)$ is the portion of $\cup C_i$ that the observer sees when looking at elevation angle $\sigma^* = \pm\tan^{-1} k$.

Each of Eqs. (3.9) and (3.10) describes a vertical cone whose line elements have slope k. The first cone has vertex at the lamp, the second has vertex at the observer's eye. The set $C^k(h)$ is the intersection of the two cones, as

shown in Fig. 3.2. The intersection is planar, and hence $C^k(h)$ is a conic section. If the cone elements are steeper than the line segment between the observer and the lamp ($k > h/2$), then $C^k(h)$ is a hyperbola on $C_1 \cup C_2$. If the cone elements are less steep ($k < h/2$), then $C^k(h)$ is an ellipse on C_0.

So Eq. (3.8) can be refined:

$$C_0 = \bigcup_{0 \le k \le h/2} C^k(h) \quad \text{(orange ellipses, Fig. 3.3)} \tag{3.11}$$

$$C_1 \cup C_2 = \bigcup_{k \ge h/2} C^k(h) \quad \text{(green hyperbolas, Fig. 3.3)} \tag{3.12}$$

Using Eqs. (3.4) and (3.5), and letting $a = h/(2k)$, you can rewrite the equations for $C^k(h)$ [Eqs. (3.9) and (3.10)] as

$$z = \frac{2k^2}{h}\, y \quad \text{(a plane)} \quad \text{and} \tag{3.13}$$

$$\begin{cases} l_1 + l_2 = 2a \quad \text{and} \quad a \ge 1 & \text{(ellipse on } C_0\text{)} \\ \quad \text{or} & \\ |l_2 - l_1| = 2a \quad \text{and} \quad a \le 1 & \text{(hyperbola on } C_1 \cup C_2\text{)} \end{cases} \tag{3.14}$$

Also see Fig. 3.4.

So all conics $C^k(h)$ with fixed $a = h/(2k)$ look the same if viewed from above. They all project to the conic with foci $(0, \pm 1)$ and with vertices $(0, \pm a)$. Its equation is

$$\frac{y^2}{a^2} + \frac{x^2}{a^2 - 1} = 1. \tag{3.15}$$

More generally, define the function a by

$$a = \frac{h}{2\,|m|}. \tag{3.16}$$

In space the level surfaces for a are the same as those for $|m|$ but with different labels. Each level surface is a vertical cone with vertex at the lamp. The value of $|m|$ on the cone gives the slope of its line elements, and the value of a gives the radius of the circle where the cone intersects the xy-plane. On the surface $\cup C_i$ the contours for a are also the same as those for $|m|$—they are the conics $C^k(h)$. The meaning of the labels $|m|$ and a are the same as they were on the cone, but now a also gives the y-coordinate of the vertices of $C^k(h)$. See Fig. 3.5.

Changing the lamp height h has a simple effect on the combined parhelic circle $\cup C_i$: if you multiply h by t then you stretch $\cup C_i$ vertically by that

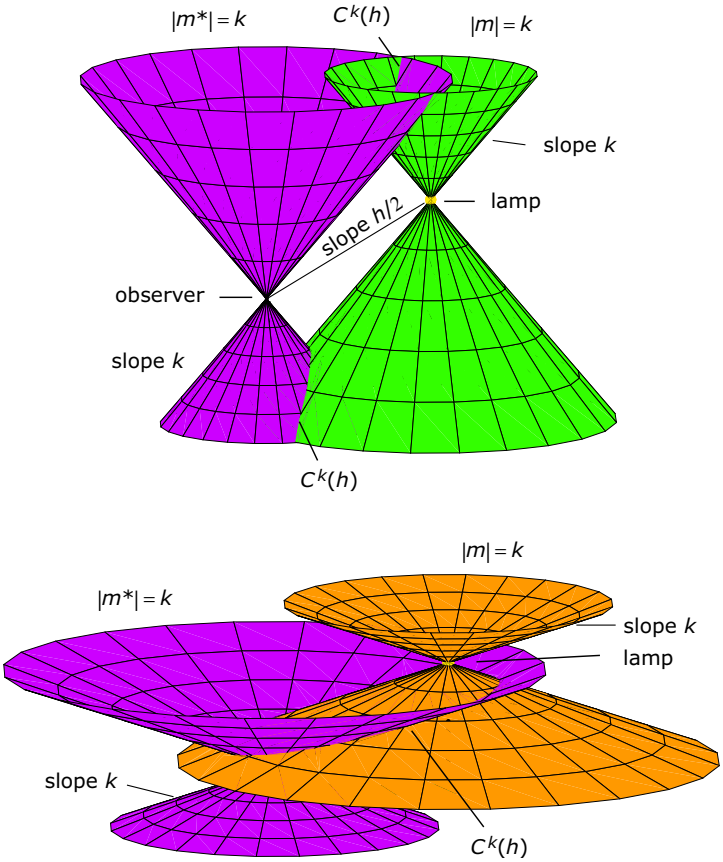

Figure 3.2: (*Top*) Cone of points (purple) with slope $|m^*| = k$, and cone of points (green) with $|m| = k$. Here $k = \tan 50^\circ > h/2$, so the intersection $C^k(h)$ of the cones is a hyperbola. (*Bottom*) Similar, but with $k = \tan 20^\circ < h/2$. Now $C^k(h)$ is an ellipse.

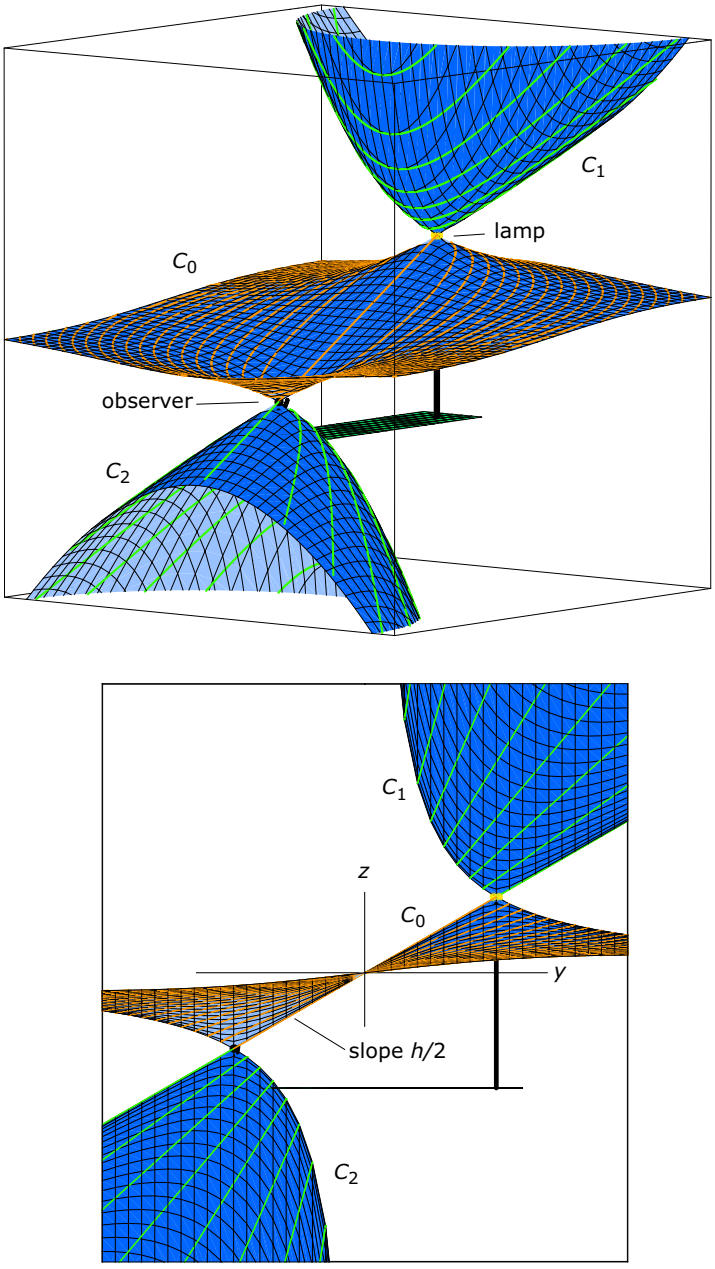

Figure 3.3: The combined parhelic circle $\cup C_i$ as a union of conics $C^k(h)$—orange ellipses on the parhelic circle C_0 and green hyperbolas on the super- and subparhelic circles $C_1 \cup C_2$. On each $C^k(h)$ the slopes $|m|$ and $|m^*|$ are constant and equal to k (Fig. 3.2). Lamp elevation $\Sigma = 30°$.

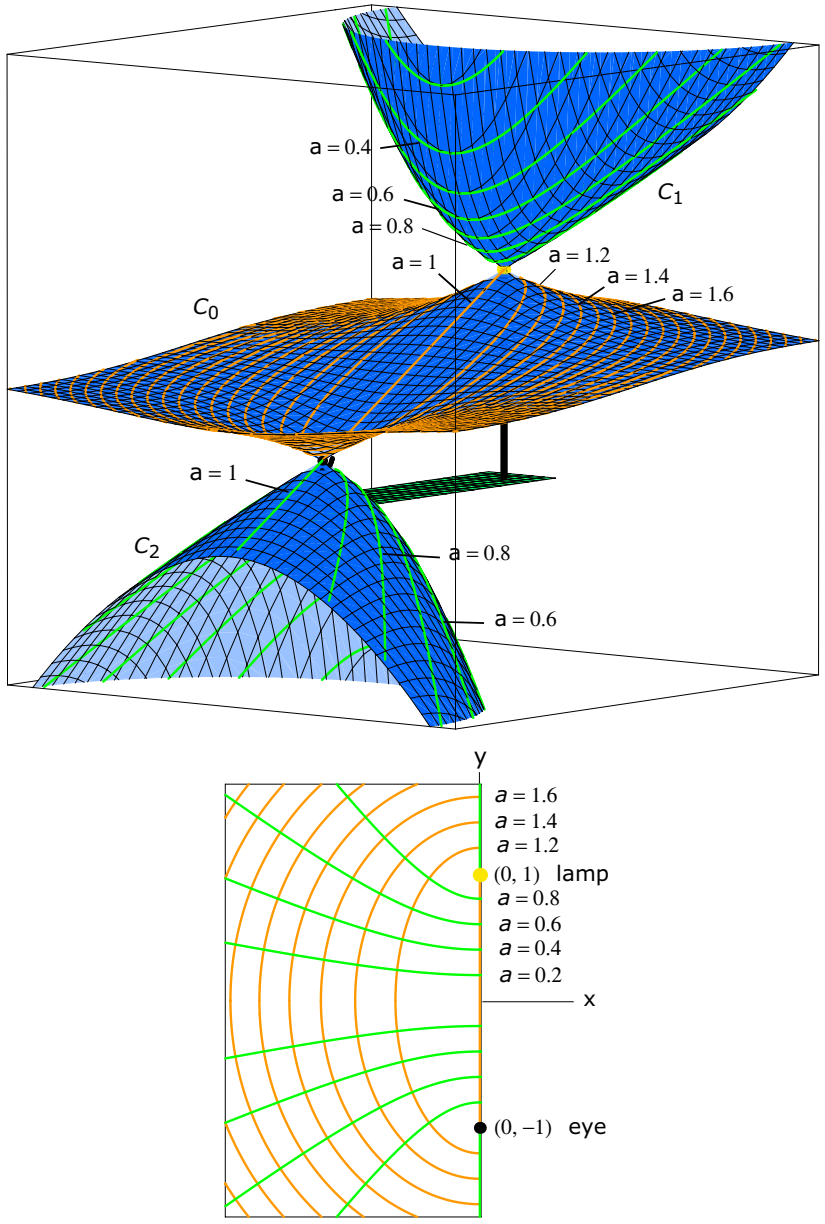

Figure 3.4: (*Top*) The combined parhelic circle as in Fig. 3.3 but with the conics $C^k(h)$ labelled with their a-values. (*Bottom*) Same but showing the left half as viewed from above. Each a-value is the y-intercept of its conic. This view is independent of lamp height h; all conics $C^k(h)$ having the same a-value look the same in projection to the xy-plane, regardless of h.

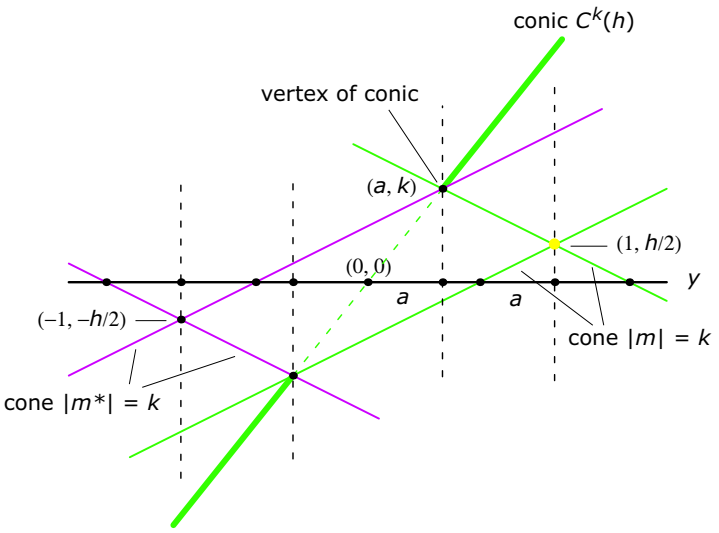

Figure 3.5: Cones $|m| = k$ (green) and $|m^*| = k$ (purple) together with their intersection, the conic $C^k(h)$ (heavy green), similar to the upper diagram in Fig. 3.2 but showing only the section in the yz-plane. The number $a = h/(2k)$ gives the horizontal coordinate of the vertex of the conic as well as the radii of the circles where the cones intersect the plane $z = 0$. Stretching the figure vertically does not change a.

same factor of t. (This is clear from Fig. 3.1. Or you can replace h and z in Eqs. (3.6) and (3.7) by $t\,h$ and $t\,z$; the equations do not change.) When you do this, you stretch the conic $C^k(h)$ to the conic $C^{t\,k}(t\,h)$, both with the same a-value. All conics $C^k(h)$ having the same a-value are stacked one above the other; they all look the same when viewed from above.

3.2 Projected parhelia

The parhelia form in horizontal hexagonal plate crystals when light rays enter one prism face of the crystal and exit an alternate prism face. The essential feature of the crystal is the idealized wedge determined by the planes of the entry and exit faces. For the parhelia the wedge axis—the line of intersection of the two wedge faces—is vertical, and the wedge angle—the angle between the faces—is 60°. The super and subparhelia are the same except that the ray also experiences an internal reflection from a basal face of the crystal. So in the wedge context you have a reflection from a plane normal to the wedge axis.

The way that I handle the parhelia, subparhelia, and superparhelia is, as already mentioned, to try first to understand their projections in the xy-plane, which is a normal plane, that is, it is perpendicular to the wedge axes. For the calculations, I will consider the left half of the halos rather than the right; we can therefore assume $x \le 0$.

3.2.1 Geometric deviation

The *geometric deviation* $d = d(x, y) = d(x, y, z)$ is the angle shown in Fig. 3.6. Analytically it is given by

$$\cos d = \frac{1 - x^2 - y^2}{l_1\, l_2} \qquad (x \le 0). \tag{3.17}$$

So d is purely geometric and has nothing to do with refractive index or wedge angle or even with light rays.

3.2.2 Physical deviation

The geometric deviation needs to be distinguished from the actual physical deviation of light that occurs when it refracts through a wedge. Let's consider lamp light that passes through a vertical wedge at (x, y, z), refracting once at entry and then again at exit. There may or may not be an intervening reflection off a horizontal face; it costs nothing to consider the super-

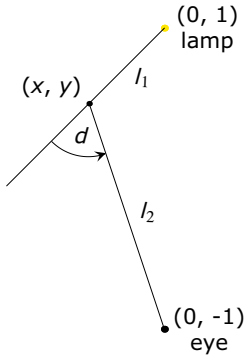

Figure 3.6: Geometric deviation angle $d(x, y)$. It depends neither on wedge angle nor refractive index. This is a view from above; the plane of the paper can be thought of as the ground.

and subparhelion $P_1 \cup P_2$ simultaneously with the parhelion P_0. Everything is to be considered as if viewed from directly above, as if projected to the xy-plane.

Then the *physical deviation* D is the angle between the entry ray and exit ray. This angle has nothing to do with the observer; we do not care, for the moment, whether the exit ray actually reaches the observer's eye.

As the wedge spins about its vertical axis—the line of intersection of the entry and exit faces—the physical deviation D varies between values $D_{\min}$ and $D_{\max}$. The angles D, $D_{\min}$, and $D_{\max}$ are given by

$$\sin(D + \alpha + \mu) = n' \sin\left(\alpha + \sin^{-1}\frac{\sin\mu}{n'}\right) \tag{3.18}$$

$$\sin\frac{D_{\min} + \alpha}{2} = n' \sin\frac{\alpha}{2} \tag{3.19}$$

$$\cos(D_{\max} + \alpha) = \cos\alpha - \sqrt{(n')^2 - 1}\sin\alpha, \tag{3.20}$$

where μ is the angle of incidence at entry, α is the wedge angle, and n' is the refractive index (but see below). Unless you are determined to do the calculations yourself, the exact form of Eqs. (3.18)–(3.20) is not crucial. The main thing is to recognize what is a function of what, namely,

$$D = D(\mu, \alpha, n') \tag{3.21}$$

$$D_{\min} = D_{\min}(\alpha, n') \tag{3.22}$$

$$D_{\max} = D_{\max}(\alpha, n'). \tag{3.23}$$

Now we can ask whether the point (x, y, z) is on the combined parhelia $\cup P_i$. That is, can the wedge at (x, y, z) be rotated about its vertical axis so that the exit ray reaches the eye of the observer? There is no hope unless (x, y, z) is on $\cup C_i$, so let's assume it to be so. That guarantees that the slope of the exit ray will be correct. The azimuth of the exit ray will be correct when

$$d = D, \tag{3.24}$$

and that in turn will be true when

$$D_{\min} \leq d \leq D_{\max}. \tag{3.25}$$

Equation (3.25) is not so easy to satisfy, as can be seen from Fig. 3.7. In that figure only the wedge location in the right-hand diagram is such that Eq. (3.25) is satisfied. But in that case Eq. (3.24) can be solved for the angle of incidence μ. Consistent with your experience of spinning prisms in sunlight, you get two solutions and thus two wedges at (x, y)—or rather at (x, y, z)—that can refract light to the eye.

Figure 3.8 gives two more wedge locations (x, y) from which the exit ray can reach the eye, but they are special. You have $d = D_{\max}$ when the ray path is tangent to the wedge at entry or exit. You have $d = D_{\min}$ when the ray path is symmetric, that is, when the 2-fold symmetry axis of the wedge is also a 2-fold symmetry axis for the ray path.

In Fig. 3.7 and Fig. 3.8 it is the light gray spray of exit rays, not the dark gray spray, that matters, since we are talking about the left parhelion.[1]

All of the above, as mentioned, is considered as if projected to the xy-plane. The lamp and eye positions are projections of the lamp and eye, the rays are projected rays, the angles d, D, $D_{\min}$, $D_{\max}$, and μ are projected angles, and the wedges in Fig. 3.7 and Fig. 3.8 are projected wedges, with the actual wedges being above or below them (i.e., in or out of the paper) on $\cup C_i$. That means that the refractive index n' is not the true refractive index but rather the Bravais refractive index

$$n' = \sqrt{n^2 + m^2(n^2 - 1)}, \tag{3.26}$$

where $n = 1.31$ is the true refractive index of ice, and where m is the slope of the entry ray.

For the classical (i.e., parallel light) parhelia and subparhelia at a fixed sun elevation the ray slope m is constant, and so n' is constant as well. In

[1]The entry face of the wedge is distinguished from the exit face, with the ray only allowed to enter the wedge at the entry face, so the dark spray cannot arise from our wedges.

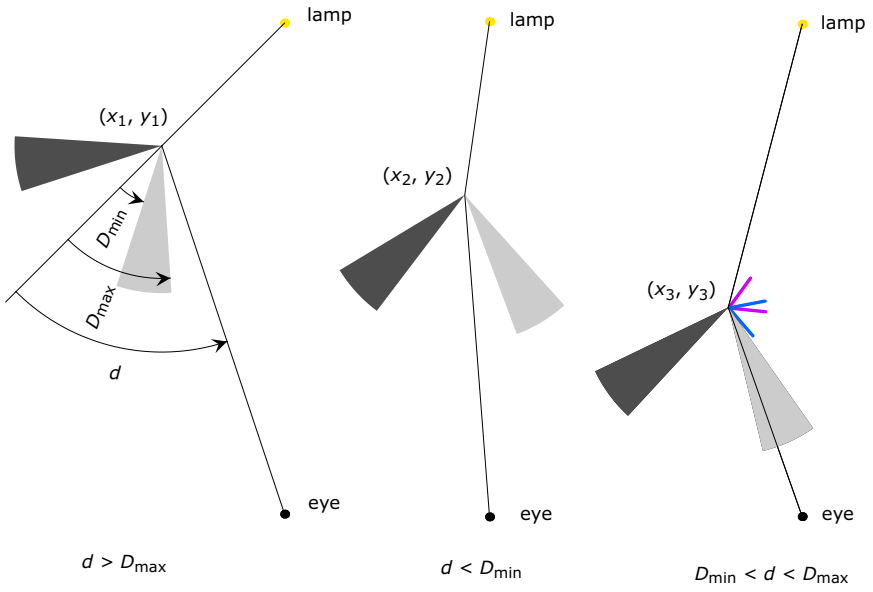

Figure 3.7: (*Left*) Sprays (gray) of exit rays from an ordinary ($\alpha = 60°, n = 1.31$) parhelion wedge (not shown) at (x_1, y_1) that is spinning about its vertical axis. The exit rays all miss the eye. (*Middle*) Same but with a different position (x_2, y_2) for the wedge. Again the exit rays miss the eye. (*Right*) A third position (x_3, y_3) for the wedge, this time giving an exit ray that reaches the eye. Two wedges (purple and blue) both give this same exit ray. The two wedges are closely related; the ray entry angle for the one is the ray exit angle for the other. These views are top views, looking straight down on the observer and the lamp. The lamp elevation is $\Sigma = 30°$, but it is not evident from the figure, due to the viewpoint. Since $\Sigma \neq 0°$, none of the rays is in the plane of the paper.

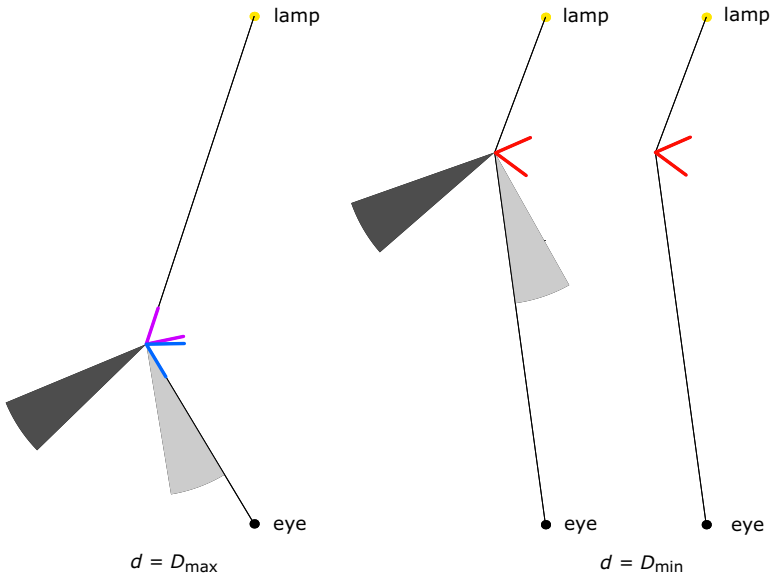

Figure 3.8: Similar to Fig. 3.7 but with two special choices for the wedge position. (*Left*) Giving $d = D_{\max}$. The exit ray that reaches the eye is the most deviated ray in the spray. The ray path is tangent to the effective wedge at entry (purple wedge) or at exit (blue wedge). (*Middle*) Giving $d = D_{\min}$. The exit ray that reaches the eye is the least deviated ray in the spray. The two effective wedges coincide to give a single wedge (red). (*Right*) Same but with the sprays removed in order to emphasize the symmetry of the ray path with respect to the wedge. The lamp elevation is $\Sigma = 30^\circ \neq 0^\circ$, so none of the rays is in the plane of the paper.

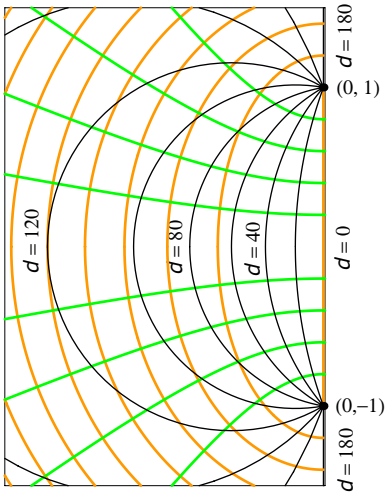

Figure 3.9: Contours (black) for the geometric deviation d together with contours for a, the latter being as in Fig. 3.4. This diagram is independent of lamp elevation angle, wedge angle, and refractive index.

our case, however, even with the lamp elevation angle Σ fixed, the ray slope varies with the position (x, y, z) of the wedge, and hence so does n'.

3.2.3 *da* coordinates

I will use the geometric deviation d and the conic vertex parameter a as coordinates on the combined parhelic circle $\cup C_i$. In projection the contours for d are arcs of circles that pass through $(0, \pm 1)$ and that have centers on the x-axis (Fig. 3.9). As explained in Section 3.1, the projected contours for a are conics with foci at $(0, \pm 1)$; they are ellipses for C_0 and hyperbolas for $C_1 \cup C_2$.

From Eqs. (3.16) and (3.26),

$$n' = n'(a, \Sigma, n) = \sqrt{n^2 + \frac{n^2 - 1}{a^2} \tan^2 \Sigma}. \tag{3.27}$$

Then using Eqs. (3.27), (3.22), (3.23), and (3.16), you can express $D_{\min}$,

$D_{\max}$, and m in terms of a, Σ, α, and n:

$$D_{\min} = D_{\min}(a, \Sigma, \alpha, n) \tag{3.28}$$
$$D_{\max} = D_{\max}(a, \Sigma, \alpha, n) \tag{3.29}$$
$$|m| = \frac{\tan \Sigma}{a} \tag{3.30}$$

The contours for a, for ray slope $|m|$, and for Bravais index n', all coincide—they are the orange ellipses on C_0 and the green hyperbolas on $C_1 \cup C_2$. You can label them with their a, $|m|$, or n' values. But labelling them with their a-values has the nice feature that the labels do not change with Σ.

In da-coordinates, Eq. (3.25), the equation for the combined parhelia $\cup P_i$, becomes

$$D_{\min}(a, \Sigma, \alpha, n) \le d \le D_{\max}(a, \Sigma, \alpha, n) \tag{3.31}$$

So for fixed a the portion of the a-contour that lies on $\cup P_i$ is the portion whose d coordinate is between the numbers $D_{\min}(a, \Sigma, \alpha, n)$ and $D_{\max}(a, \Sigma, \alpha, n)$, both of which are constant on the contour. And the boundary of $\cup P_i$ consists of the curves

$$d = D_{\min}(a, \Sigma, \alpha, n) \tag{3.32}$$
$$d = D_{\max}(a, \Sigma, \alpha, n). \tag{3.33}$$

Figure 3.10 shows the resulting combined parhelia $\cup P_i$ for a lamp elevation $\Sigma = 30°$. The view is from above, with everything seen as if projected to the horizontal plane. Then Fig. 3.11, which is the corresponding spatial view, was made from the projected view by using Eqs. (3.6) and (3.7) to move it up or down to $\cup C_i$. Rather, that gave half the figure; the other half was added by symmetry.

3.2.4 Wedges

As mentioned previously, at a typical point of the parhelion there are two wedges that refract lamp light to the eye. Wedges on the $\Sigma = 30°$ parhelion are shown in Fig. 3.13.

3.2.5 da versus xy

I have shown how to calculate $D_{\min}(a, \Sigma, \alpha, n)$ and $D_{\max}(a, \Sigma, \alpha, n)$, thus making it easy to plot the curves $d = D_{\min}(a, \Sigma, \alpha, n)$ and $d = D_{\max}(a, \Sigma, \alpha, n)$

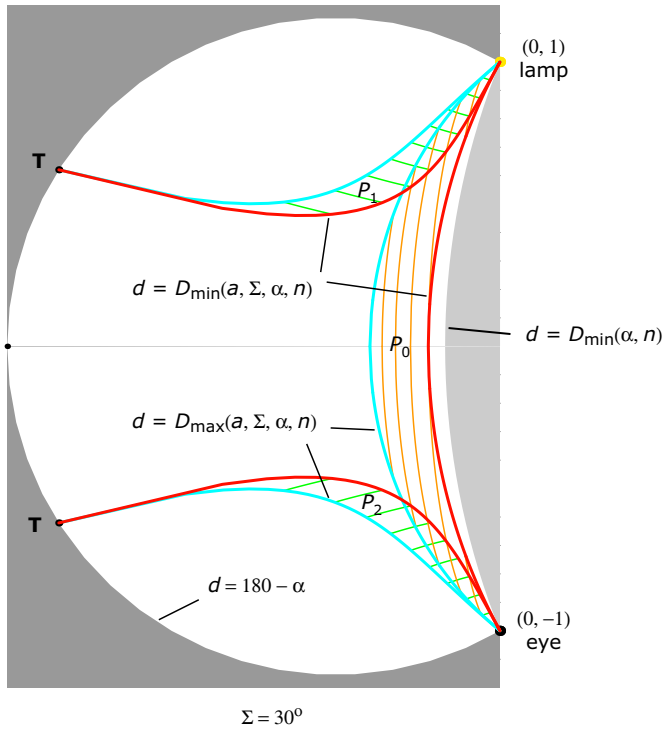

Figure 3.10: Ordinary ($\alpha = 60°$, $n = 1.31$) left parhelion P_0, superparhelion P_1, and subparhelion P_2 for a lamp elevation of $\Sigma = 30°$. They make up the region between the red curve $d = D_{\min}(a, \Sigma, \alpha, n)$ and the turquoise curve $d = D_{\max}(a, \Sigma, \alpha, n)$. The orange elliptical arcs are a-contours on P_0, and the green hyperbolic arcs are a-contours on $P_1 \cup P_2$; the Bravais refractive index n' is constant on each. Everything is seen from above, looking directly down on the observer and the lamp, and it is not evident that P_0, P_1, P_2 are on separate surfaces (C_0, C_1, C_2).

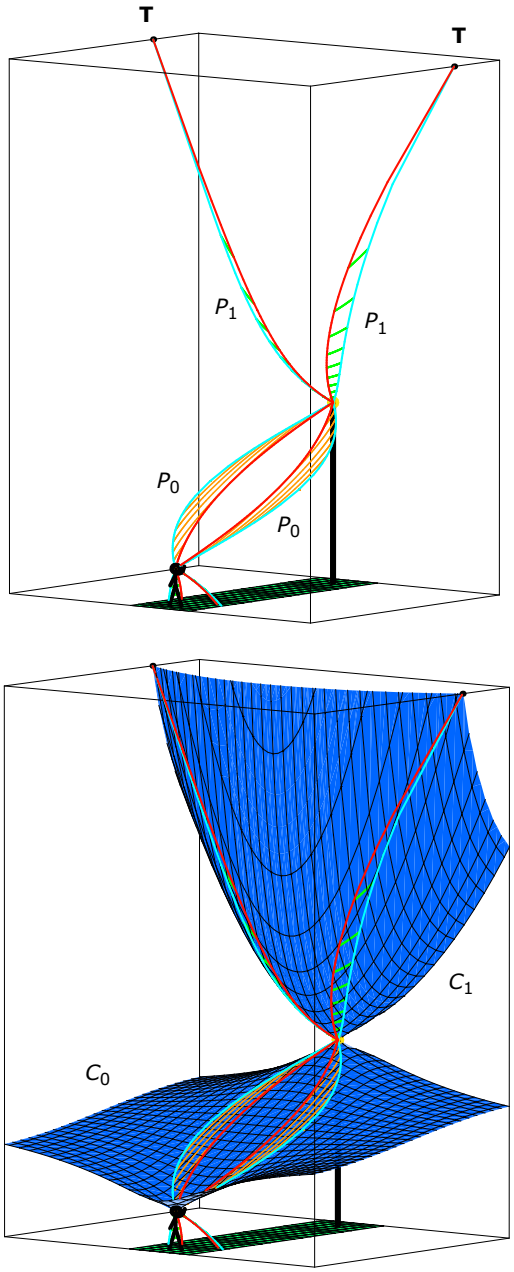

Figure 3.11: (*Top*) Parhelia P_0, superparhelia P_1, and a small piece of the subparhelia for a lamp elevation of $\Sigma = 30°$. Compare with Fig. 3.10, which is a top view of the left-hand side of these same halos. (*Bottom*) Same but with the addition of the parhelic circle C_0 and superparhelic circle C_1.

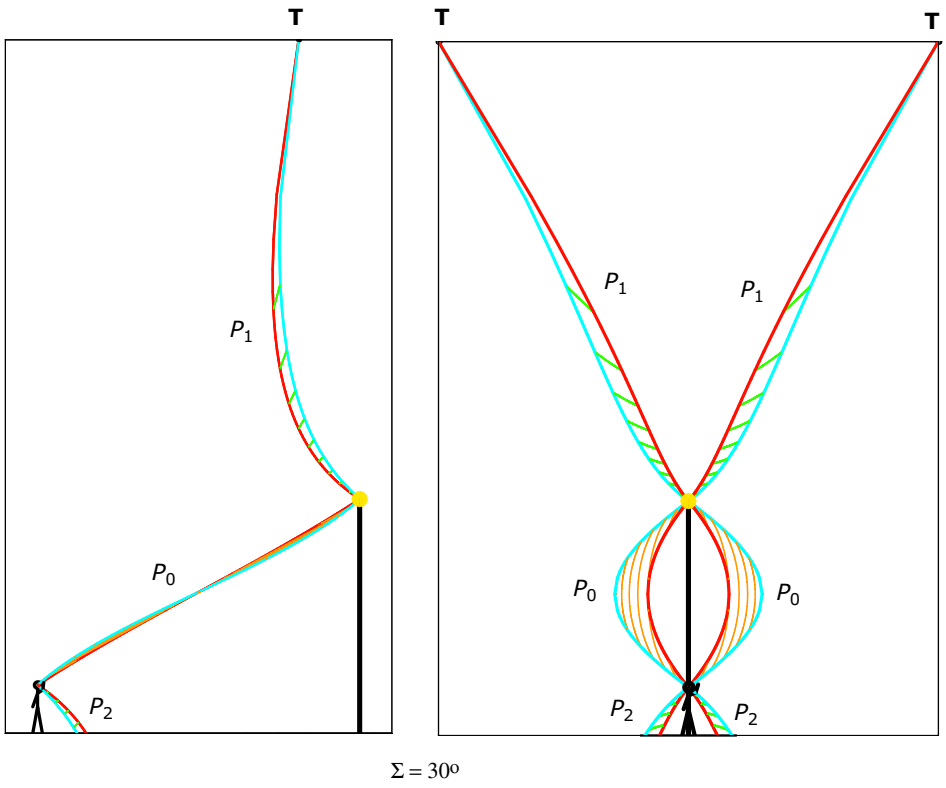

Figure 3.12: Side and end views of the combined parhelia of Fig. 3.11.

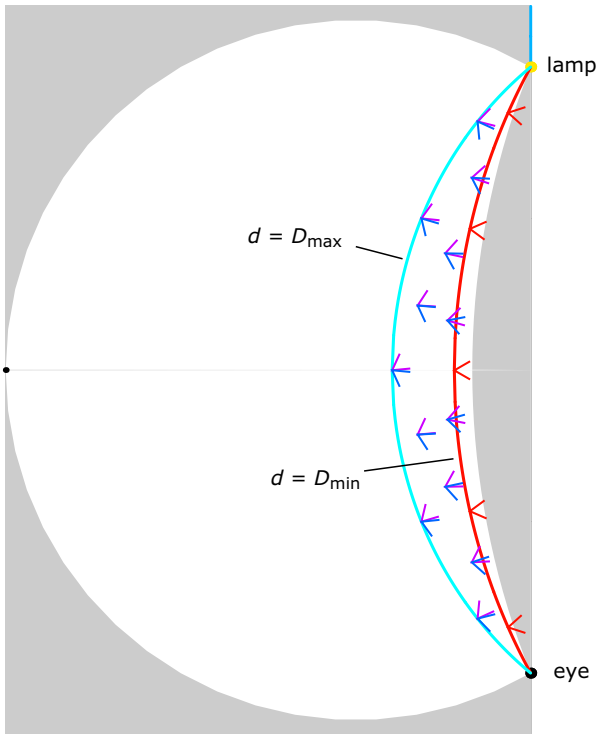

Figure 3.13: The parhelion of Fig. 3.10 together with some of the halo-making wedges. At each point on the boundary curve $d = D_{\max}$ the ray path (not shown) would be tangent to the wedge at entry or exit. At each point on the boundary curve $d = D_{\min}$—the caustic of the parhelion—the ray path would be symmetric. See also Fig. 3.8.

in the da-plane. But in the figures they were plotted on $\cup C_i$ in xyz-space (Fig. 3.11) or else in projection on the xy-plane (Fig. 3.10). How did that happen? What got swept under the rug was the relation between the da coordinates and the xy coordinates.

Equations (3.14) and (3.17) take xy coordinates to da coordinates. In the other direction, the mapping taking da coordinates to xy coordinates on C_0 is

$$(x,\, y) = \left(-\frac{(a^2-1)\cos(d/2)}{\sin(d/2)},\, \frac{a\sqrt{1-a^2\cos^2(d/2)}}{\sin(d/2)} \right) \qquad \text{if } 1 \le a \le \sec(d/2), \tag{3.34}$$

and the mapping taking da coordinates to xy coordinates on $C_1 \cup C_2$ is

$$(x,\, y) = \left(-\frac{(1-a^2)\sin(d/2)}{\cos(d/2)},\, \frac{a\sqrt{1-a^2\sin^2(d/2)}}{\cos(d/2)} \right) \qquad \text{if } 0 < a \le 1. \tag{3.35}$$

Equations (3.34) and (3.35) give $x \le 0$, $y \ge 0$. The other quadrants follow from symmetry.

Figure 3.14 illustrates the relation between the da coordinates and the xy coordinates. Figure 3.15 is similar but with the addition of the parhelion P_0 and superparhelion P_1.

3.3 Dependence on lamp elevation

Figures 3.10, 3.11, 3.16–3.18, and 3.23 show the evolution of the combined parhelia $\cup P_i$ with lamp elevation Σ. In this section I will describe one way to think about this evolution.

There are three broad constraints on $\cup P_i$. They do not characterize $\cup P_i$ completely, but they give some insight into why it evolves the way it does. First, of course, $\cup P_i$ must lie on the combined parhelic circle $\cup C_i$. Second, for any point on $\cup P_i$,

$$D_{\min}(\alpha, n) \le D_{\min}(a, \Sigma, \alpha, n) \le\ d \le D_{\max}(a, \Sigma, \alpha, n) \le 180 - \alpha. \tag{3.36}$$

On the left of the equation, the number $D_{\min}(\alpha, n)$ is the familiar minimum deviation for the circular halo, that is, it is the deviation for a ray path that is symmetric and that lies in the normal plane. On the far right the number $180 - \alpha$ is the (projected) deviation when the ray path is doubly tangential, that is, when it is tangent to the wedge at both entry and exit.

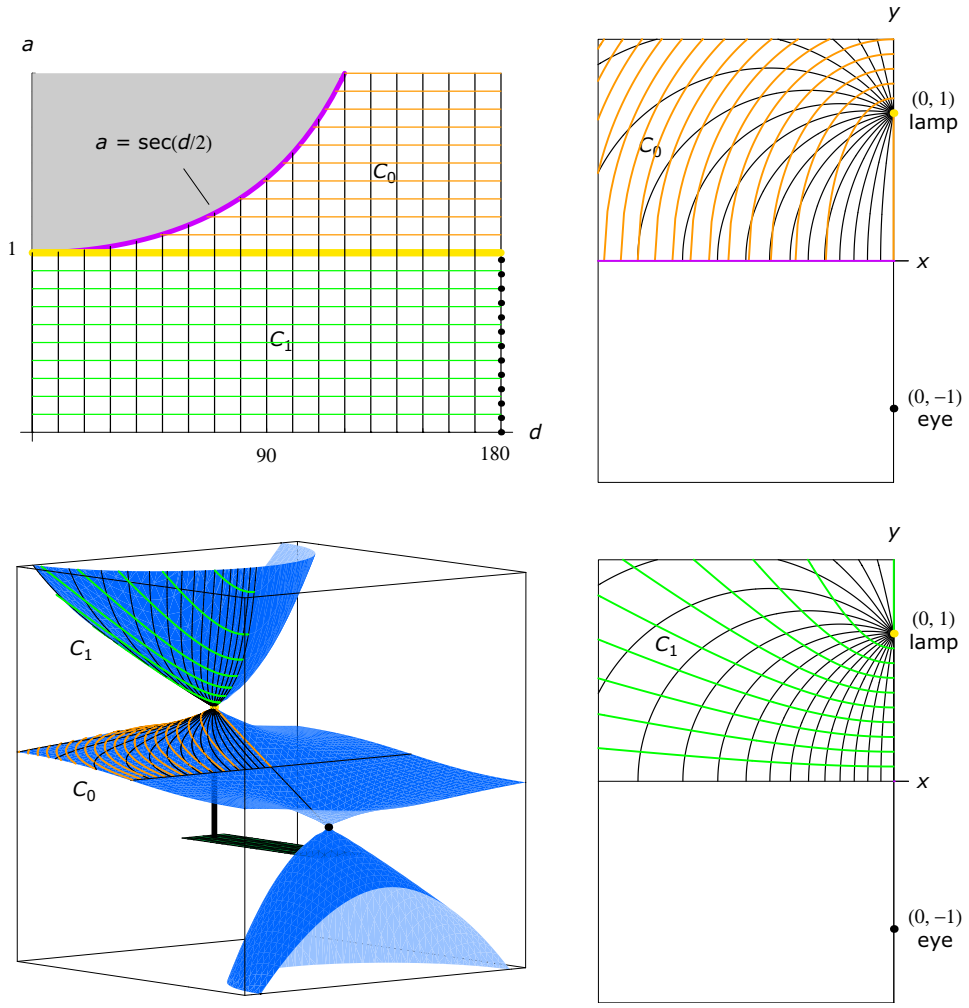

Figure 3.14: Relation between da coordinates and xy coordinates on $\cup C_i$. The black vertical lines (d = constant) in the da-plane correspond to the black circular arcs in the xy-plane. The orange horizontal lines (a = constant > 1) correspond to the orange elliptical arcs, and the green horizontal lines (a = constant < 1) correspond to the green hyperbolic arcs. The purple curve $a = \sec(d/2)$ corresponds to the x-axis. The entire yellow segment $a = 1$ corresponds to the yellow point $(0, 1)$ at the lamp. Here $x \leq 0$ and $y \geq 0$; the other quadrants follow from symmetry. The lamp elevation is $\Sigma = 30°$, but only the lower left diagram is affected by Σ.

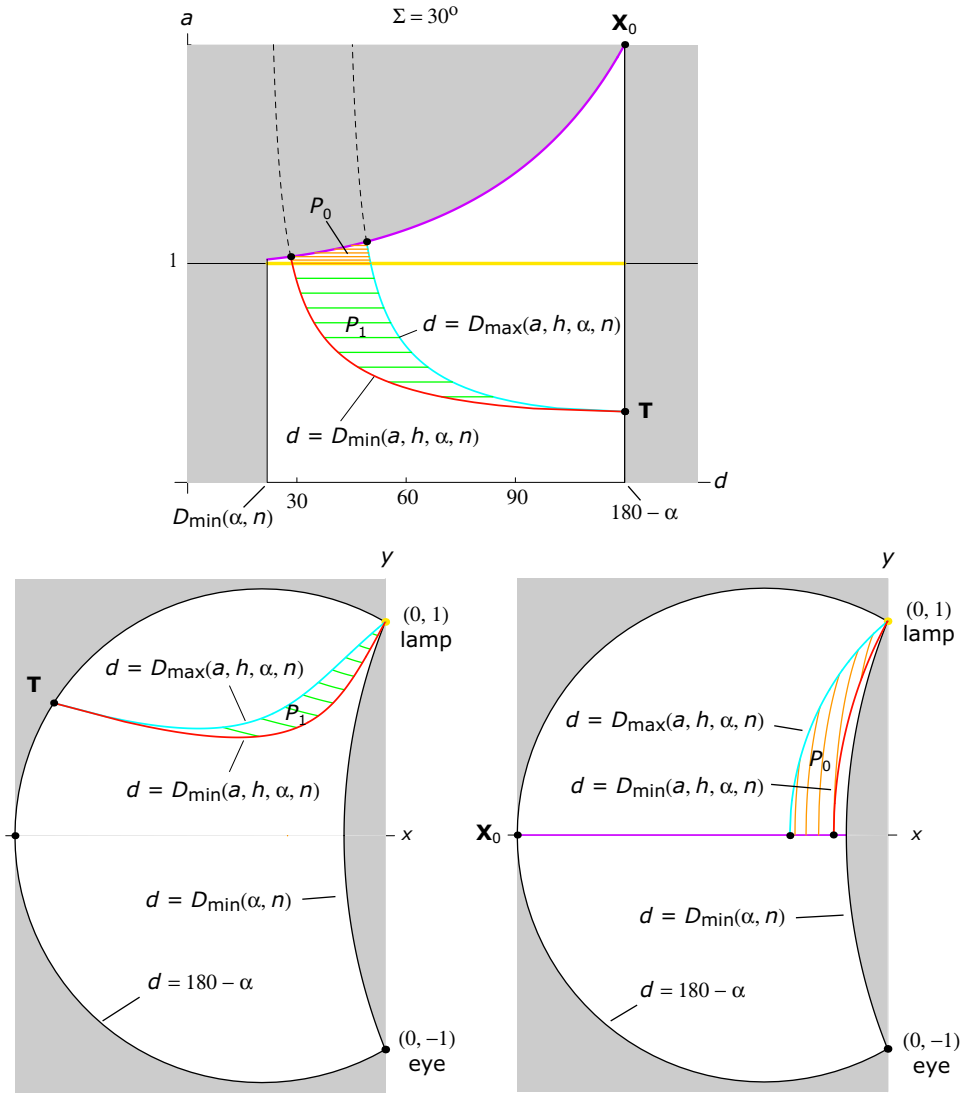

Figure 3.15: Left superparhelion P_1 and half of the left parhelion P_0 for $\Sigma = 30°$, shown in both the da-plane and the xy-plane. Like-colored curves correspond to each other, as explained in Fig. 3.14. Figure 3.10 comes from this figure by superposing the two lower diagrams and using symmetry. These diagrams are for $\alpha = 60°$ and $n = 1.31$ as usual, so $D_{\min}(\alpha, n) = 21.8°$ and $180 - \alpha = 120°$.

Since those left- and right-hand numbers are independent of lamp elevation Σ, then $\cup P_i$ is confined to the region $D_{\min}(\alpha, n) \le d \le 180 - \alpha$, regardless of Σ. In projection, this is the white region in Fig. 3.10.

Finally, whether or not an exit ray reaches the eye, a ray path is doubly tangential when $D_{\min}(\alpha, n') = D_{\max}(\alpha, n') = 180 - \alpha$ or, equivalently, $n' = \csc(\alpha/2)$ [from Eq. (3.19)]. Larger n' do not permit lamp light to pass through the wedge, regardless of how the wedge spins about its vertical axis.[2] Using Eqs. (3.26) and (3.16), we can define the values for m and a that give double tangency:

$$m_0(\alpha, n) = \sqrt{\frac{\csc^2(\alpha/2) - n^2}{n^2 - 1}} \tag{3.37}$$

$$a_0(\Sigma, \alpha, n) = \frac{\tan \Sigma}{m_0}. \tag{3.38}$$

Then the conditions $|m| = m_0(\alpha, n)$ and $a = a_0(\Sigma, \alpha, n)$ both define the same cone in space, the *double tangency cone*, and they define the same conic, the *double tangency conic*, on $\cup C_i$. On the double tangency cone our spinning wedges permit only the doubly tangent ray path, and inside the cone the spinning wedges allow no ray paths at all. So $\cup P_i$ is always on or outside the double tangency cone.

Taking all three constraints into account, you find that P_i is always confined to the subset B_i of C_i defined by

$$D_{\min}(\alpha, n) \le d \le 180 - \alpha \tag{3.39}$$

$$a_0(\Sigma, \alpha, n) \le a. \tag{3.40}$$

In projection the set $\cup B_i$ is simple, and you could draw it yourself, with the help of the coordinate contours of Fig. 3.9. Figure 3.19, for example, shows $\cup B_i$ when $a_0 = 0.2$. The boundaries of $\cup B_i$ are formed by the circular arcs $d = D_{\min}(\alpha, n)$ and $d = 180 - \alpha$ and by the conic $a = a_0$. Only the conic depends on Σ, so it is easy to see how $\cup B_i$ changes with increasing Σ or, equivalently, with increasing a_0. The set $B_1 \cup B_2$ shrinks continuously, up until $a_0 = 1$, at which time $B_1 \cup B_2$ is reduced to just the two points at the lamp and the eye. For larger a_0, the set $B_1 \cup B_2$ disappears, B_0 is detached from the eye and lamp, and B_0 shrinks. By the time that $a_0 = \csc(\alpha/2)$ the set B_0 is reduced to the single point $\mathbf{X}_0$ (two points if you count both left and right), and then B_0 disappears for larger a_0. See Fig. 3.20.

[2] Always we are talking about ray paths that enter and exit the (vertical) wedge at the prescribed entry and exit faces and that perhaps undergo a horizontal reflection as well. No other reflections are allowed.

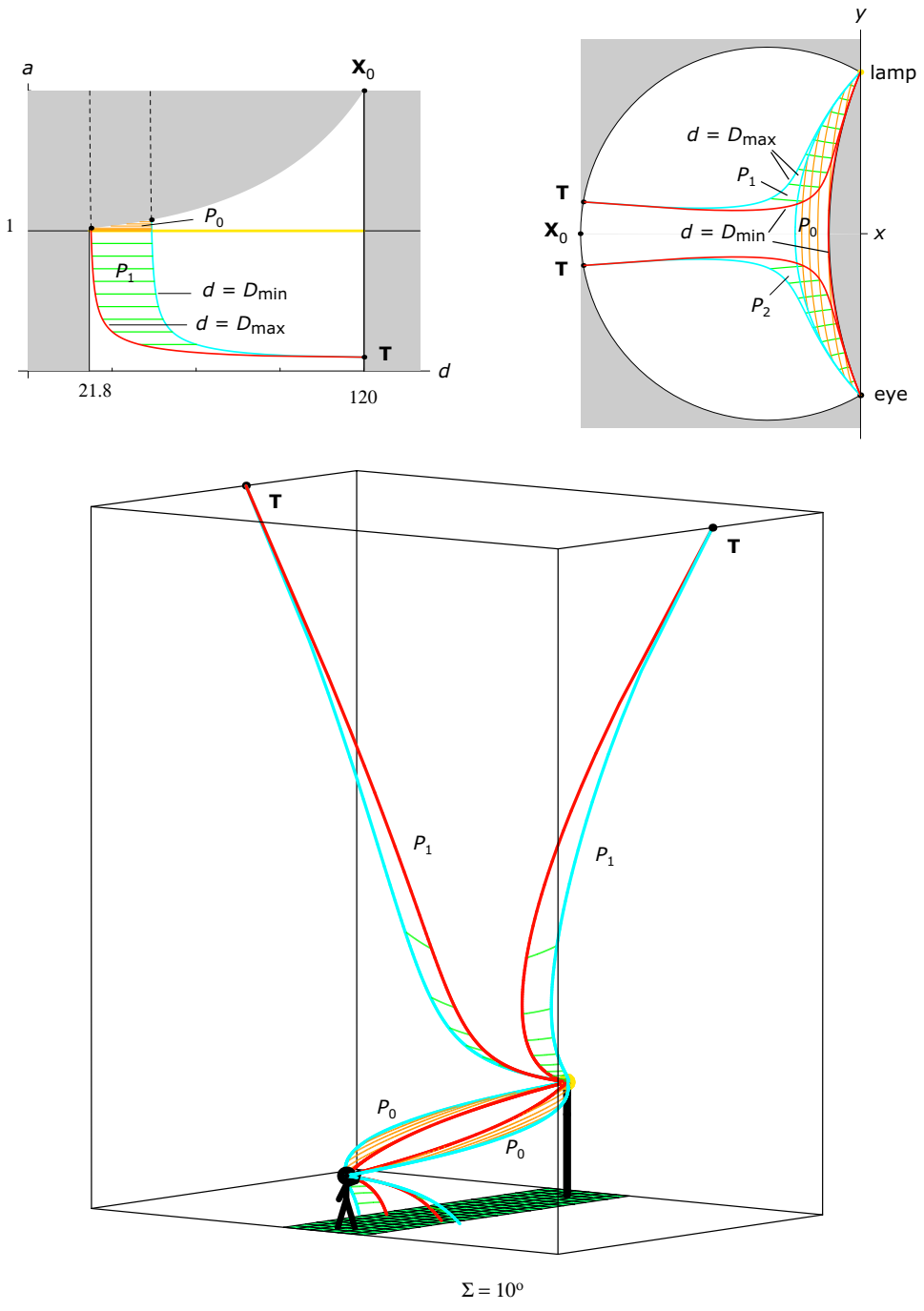

Figure 3.16: Combined parhelia for $\Sigma = 10°$.

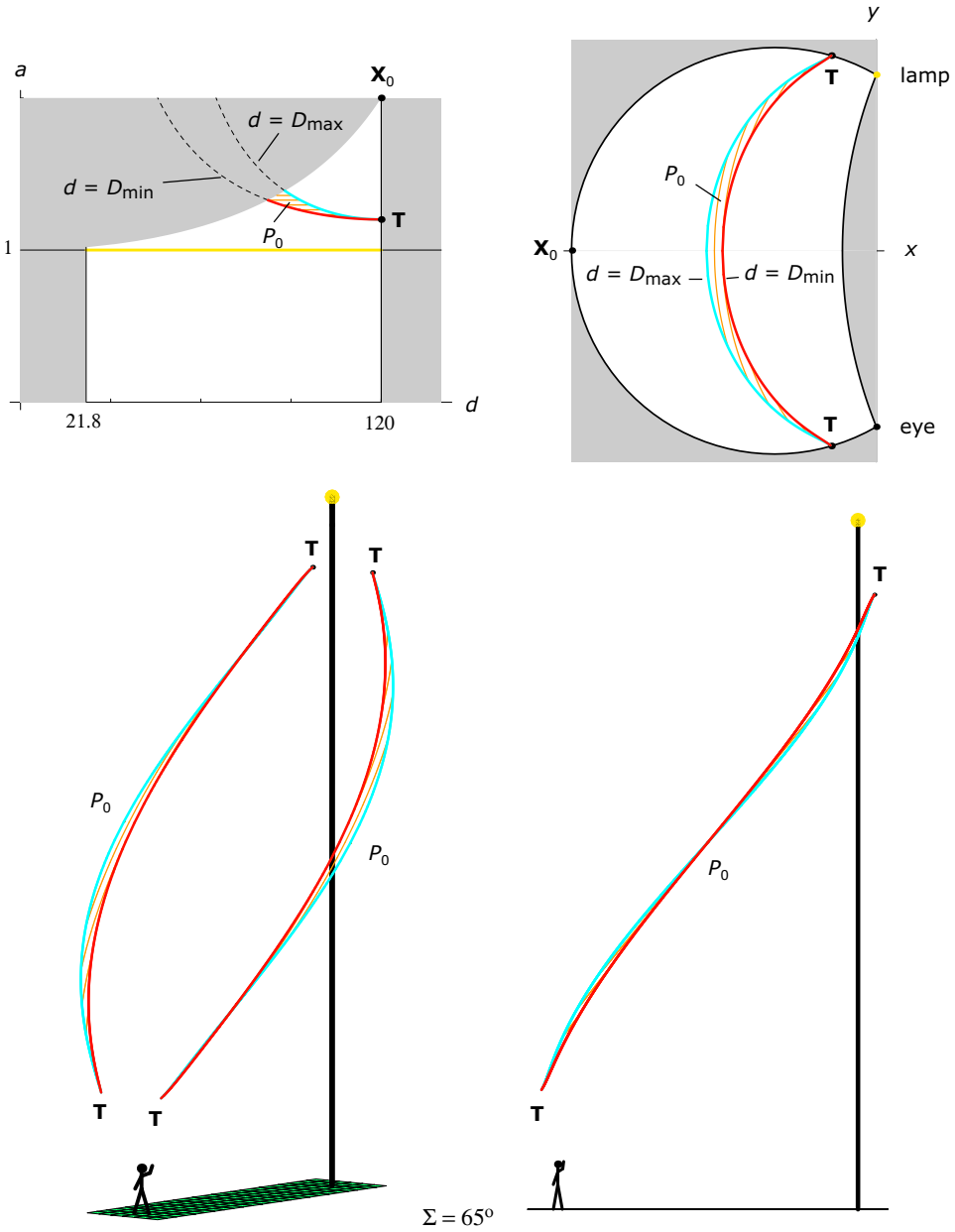

Figure 3.17: Parhelia P_0 for lamp elevation $\Sigma = 65°$. The super- and sub-parhelia have disappeared, and the parhelia are detached from the lamp and eye.

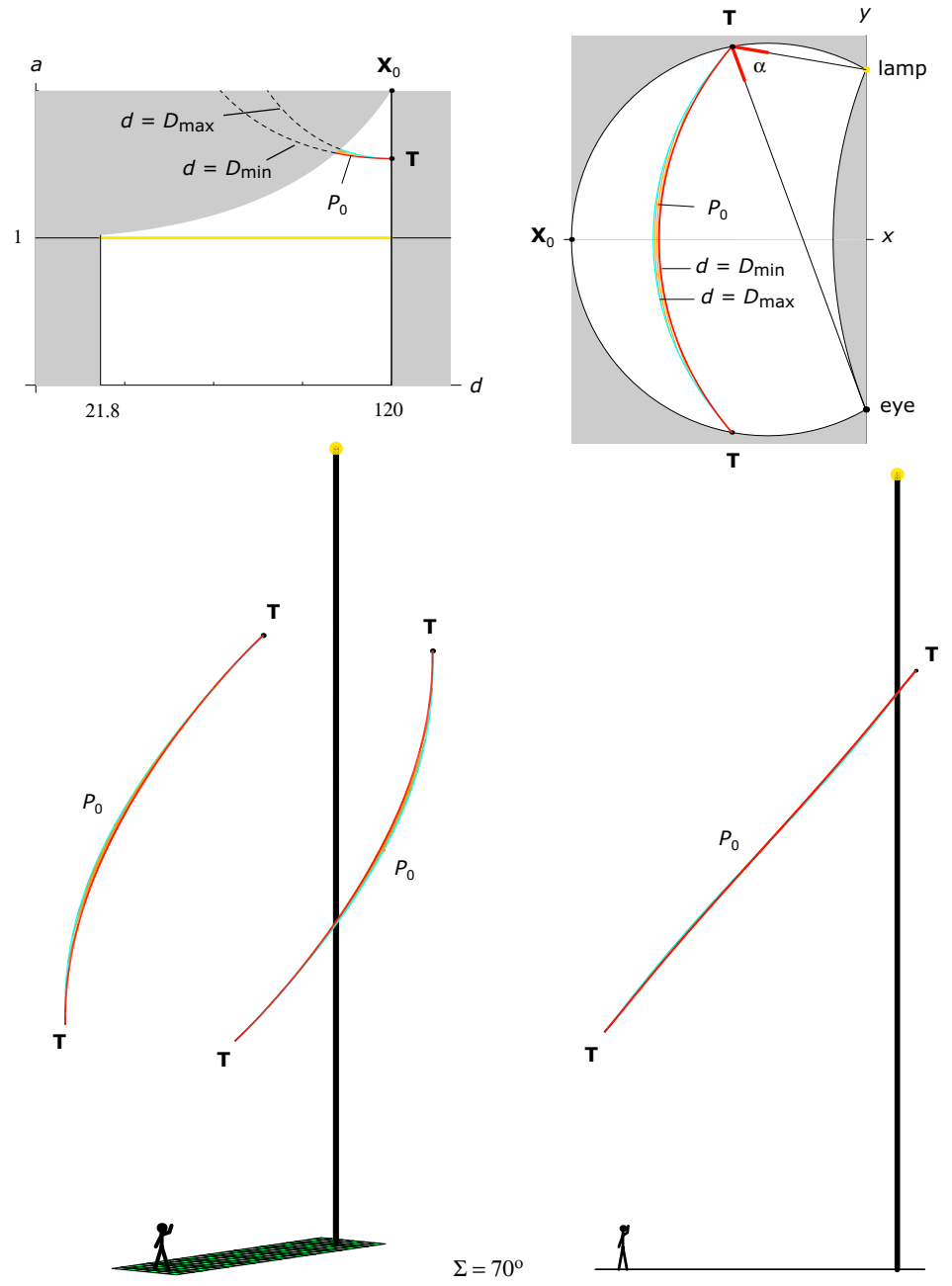

Figure 3.18: Parhelia P_0 for lamp elevation $\Sigma = 70°$. In the upper right diagram the halo-making wedge (red) has been added at the parhelion tip **T** in order to illustrate the behavior there; the ray path is doubly tangential, and $D_{\min} = D_{\max} = 180° - \alpha$.

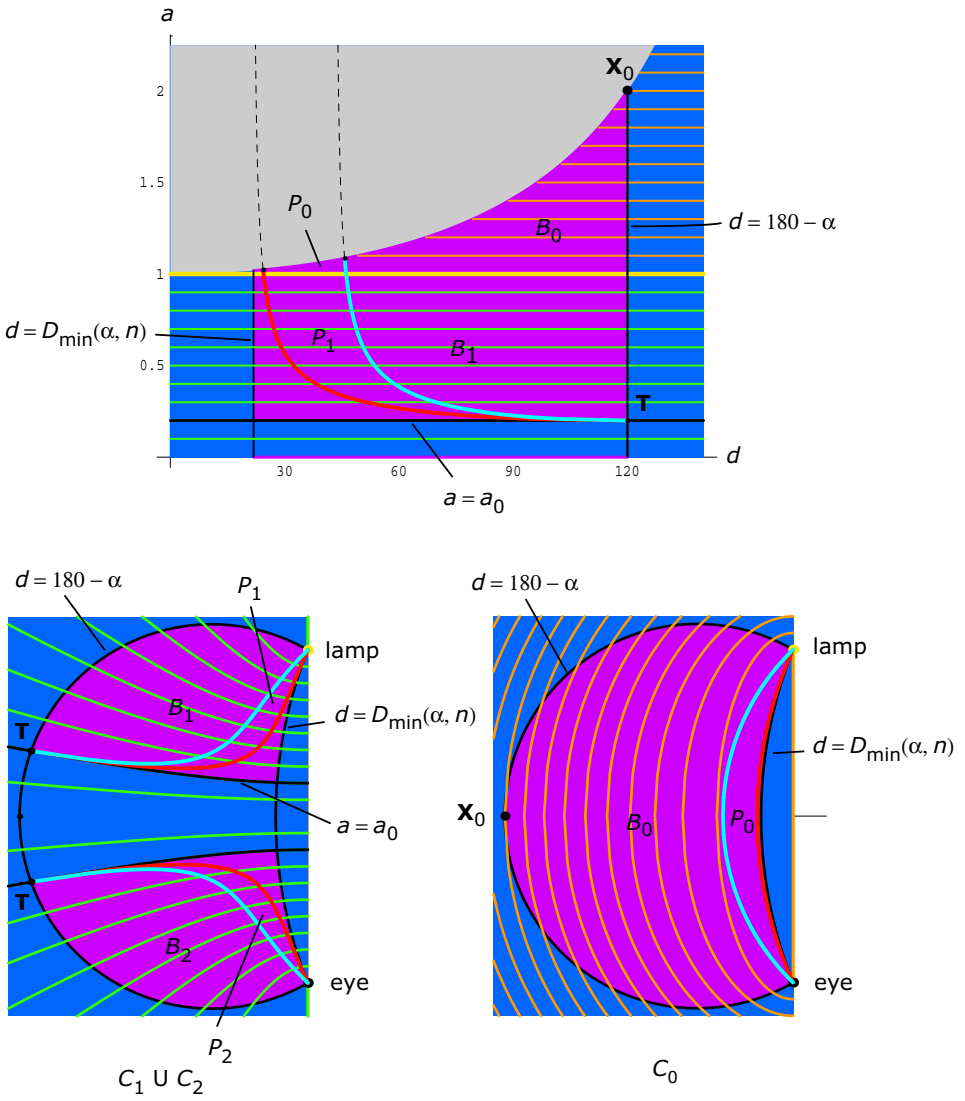

Figure 3.19: The region $\cup B_i$ (purple), to which the combined parhelia $\cup P_i$ are confined. The boundaries of $\cup B_i$ are formed by $d = D_{\min}(\alpha, n)$, $d = 180 - \alpha$, and $a = a_0$, with only the boundary $a = a_0$ depending on lamp elevation. Here $a_0 = 0.2$, corresponding to a lamp elevation of $\Sigma = 19.7°$.

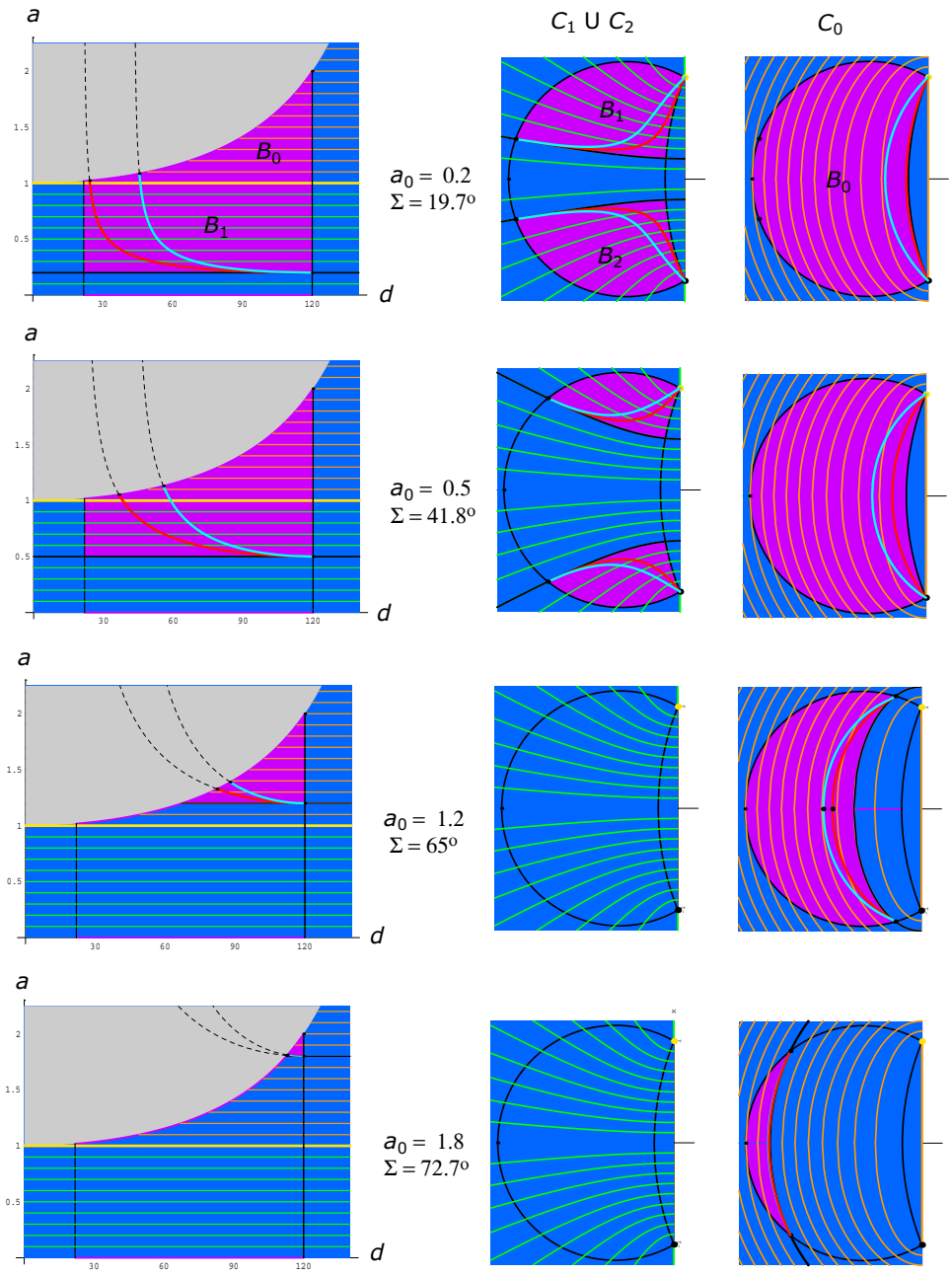

Figure 3.20: Shrinking of $\cup B_i$ (purple) with increasing Σ. The top row has $\Sigma = 19.7°$, the same as in Fig. 3.19.

During the evolution just described, the combined parhelia always lie within $\cup B_i$. When $B_1 \cup B_2$ is reduced to the lamp and the eye, so is $P_1 \cup P_2$, and when B_0 is reduced to $\mathbf{X}_0$ so is P_0. By the way, the corner $\mathbf{T}$ of $\cup B_i$, where $a = a_0$ and $d = 180 - \alpha$, is always on P and in fact is the tip of P.

For the lamp elevations $\Sigma_1(\alpha, n)$ and $\Sigma_2(\alpha, n)$ corresponding to $a_0 = 1$ and $a_0 = \csc(\alpha/2)$, respectively, you find from Eq. (3.38) that

$$\tan \Sigma_1 = m_0 \quad (P_1 \cup P_2 \text{ disappears, } P_0 \text{ detaches}) \tag{3.41}$$

$$\tan \Sigma_2 = m_0 \csc(\alpha/2) \quad (P_0 \text{ disappears}). \tag{3.42}$$

For the common case $\alpha = 60^\circ$ and $n = 1.31$, Eq. (3.41) gives $\Sigma_1 = 60.8^\circ$; that is the lamp elevation at which the parhelia detach from the lamp and eye, and at which the super- and subparhelia disappear. And Eq. (3.42) gives $\Sigma_2 = 74.4^\circ$; that's when the parhelia disappear.

We now know how the combined parhelia evolve with lamp elevation, but I am going to restate things anyway, more conceptually and in space, so that the role of the double tangency cone becomes clear. Figures 3.21 and 3.22 show the two main possibilities, namely, $0 < \Sigma < \Sigma_1$ and $\Sigma_1 < \Sigma < \Sigma_2$.

As Σ increases, starting from zero, $a_0(\Sigma, \alpha, n)$ increases as well, and the double tangency conic $a = a_0$, which is a hyperbola on $C_1 \cup C_2$, moves outward. As seen from above, B_0 does not change but $B_1 \cup B_2$ shrinks, since $a = a_0$ is a part of its boundary. At lamp elevation $\Sigma = \Sigma_1$, the double tangency conic reaches the lamp and eye, and $B_1 \cup B_2$ is reduced to just two points—the lamp and the eye. For larger Σ, and hence larger $a_0(\Sigma, \alpha, n)$, the double tangency conic drops onto C_0, becoming an ellipse; now $B_1 \cup B_2$ is empty, and B_0 is detached from the lamp and eye. As Σ continues to increase, the double tangency conic moves outward on C_0, shrinking B_0. Eventually, at $\Sigma = \Sigma_2$, the double tangency conic reaches the point $\mathbf{X}_0 = (-\cot(\alpha/2),\, 0,\, 0)$, and B_0 contracts to $\mathbf{X}_0$. For larger Σ the set B_0 is empty.

The spatial approach gives a derivation of the equations for Σ_1 and Σ_2 that is more conceptual than simply equating a_0 to 1 and $\csc(\alpha/2)$, as was done above. At $\Sigma = \Sigma_1$ the double tangency cone passes through the eye. At that moment the line segment from the eye to the lamp is therefore on the cone, making its slope equal to m_0. But that segment always has slope $\tan \Sigma$; equating the two slopes gives Eq. (3.41). And at $\Sigma = \Sigma_2$ the double tangency cone passes through the point $\mathbf{X}_0$, so now the line segment from $\mathbf{X}_0$ to the lamp has slope m_0. The horizontal distance from $\mathbf{X}_0$ to the lamp is $\csc(\alpha/2)$, and the vertical distance is $\tan \Sigma$; equating the slopes gives Eq. (3.42).

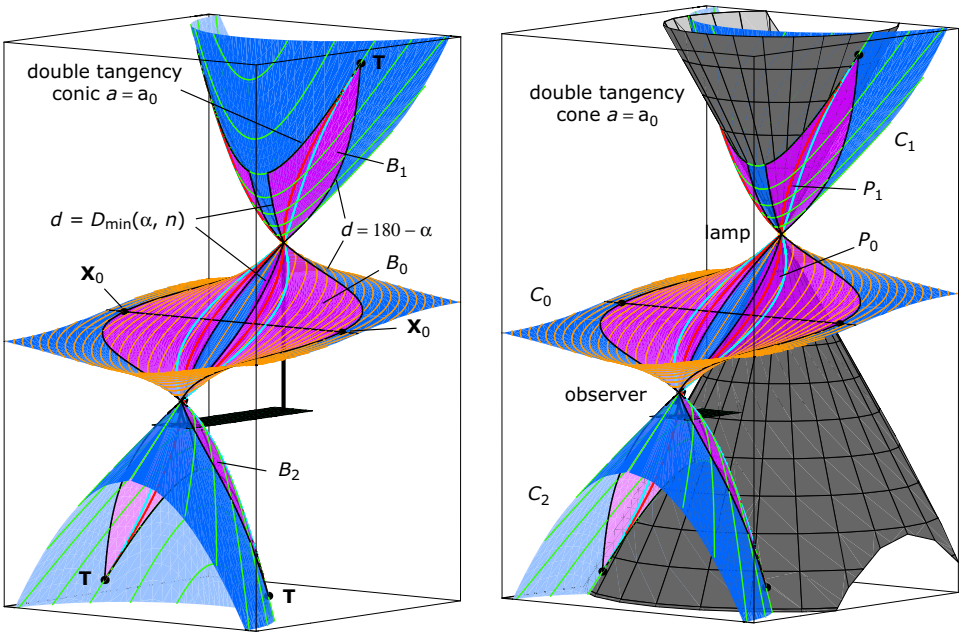

Figure 3.21: The surface $\cup B_i$ (purple), shown with and without the double tangency cone, which determines one of the boundary curves $a = a_0$ of $\cup B_i$. Inside the double tangency cone, lamp light cannot refract through vertical wedges, because the rays are too steep. The lamp elevation here is $\Sigma = 41.8°$ ($a_0 = 0.5$), the same as the second row of Fig. 3.20.

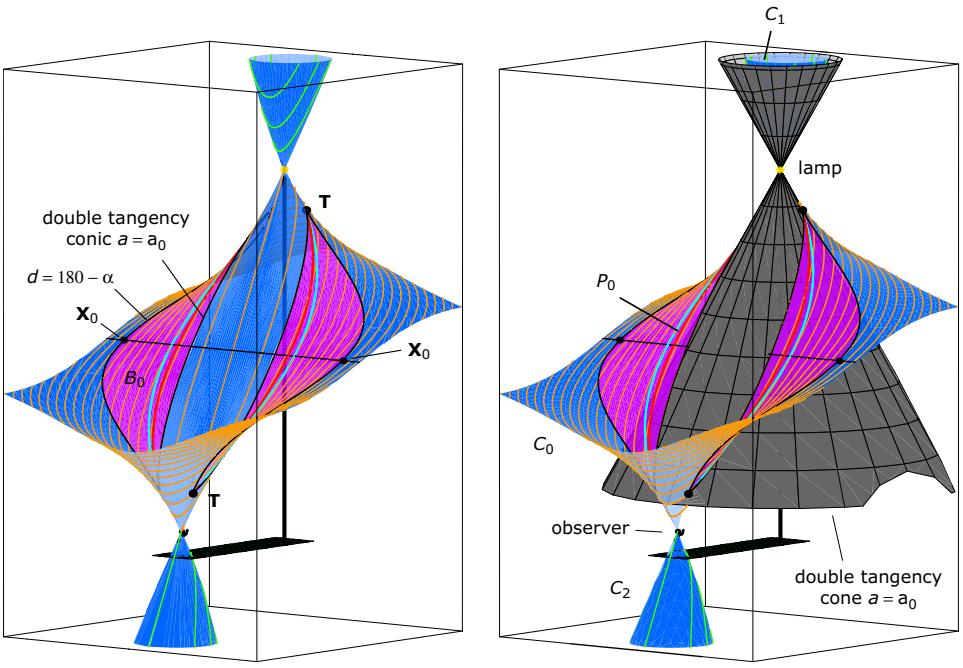

Figure 3.22: The set $\cup B_i$ (purple) for a lamp elevation of $\Sigma = 65°$ ($a_0 = 1.2$), the same as the third row of Fig. 3.20. Since $\Sigma > \Sigma_1$, the set $B_1 \cup B_2$, and hence $P_1 \cup P_2$, is now empty. And B_0, and hence P_0, is detached from the lamp and eye. The shape of the double tangency cone is the same as in Fig. 3.21; the cone's vertex moves with the lamp, but its shape does not change.

Lamp elevation $\Sigma = 0$

For the parhelion the case $\Sigma = 0$, where the lamp is at the observer's eye level, is simple. The parhelic circle C_0 becomes the xy-plane—the horizontal plane through the eye and the lamp. The parhelia lie in that plane, and everywhere in the plane the Bravais refractive index n' is constant and equal to the true refractive index n. The parhelia P_0 occupy the region between the circular arcs $d = D_{\min}(\alpha, n)$ and $d = D_{\max}(\alpha, n)$, and so they can be easily found, using none of the mathematics developed above. They are shown in Fig. 3.23, where they have been colored in their entirety, as a reminder that n' is constant there.

The super- and subparhelia, however, may be a bit of a surprise. For $\Sigma = 0$ the super- and subparhelic circles do not just coincide with the xy-plane, as you might expect from your experience with the classical halos. They also include the xz-plane, and there are indeed superparhelia and subparhelia lying on the xz-plane, as illustrated in Fig. 3.23.

Reminder

I said earlier that this chapter would not say much about how the halos would look to the observer, and it doesn't. It just calculates the loci of halo-making wedges that send lamp light to the eye. These loci are only the first step in deciding what the halos really look like.

So, for example, although I dismissed the combined parhelia for $\Sigma < 0$ as following from those for $\Sigma > 0$, in fact the two cases in practice look much different to the observer.

Another example: The high lamp elevations such as those in Fig. 3.17 and Fig. 3.18, though interesting theoretically, are apt to be irrelevant in practice, due to the highly skew ray paths involved. And any part of a halo where the deviation is near $D_{\max}$ will also probably be too weak to be seen.

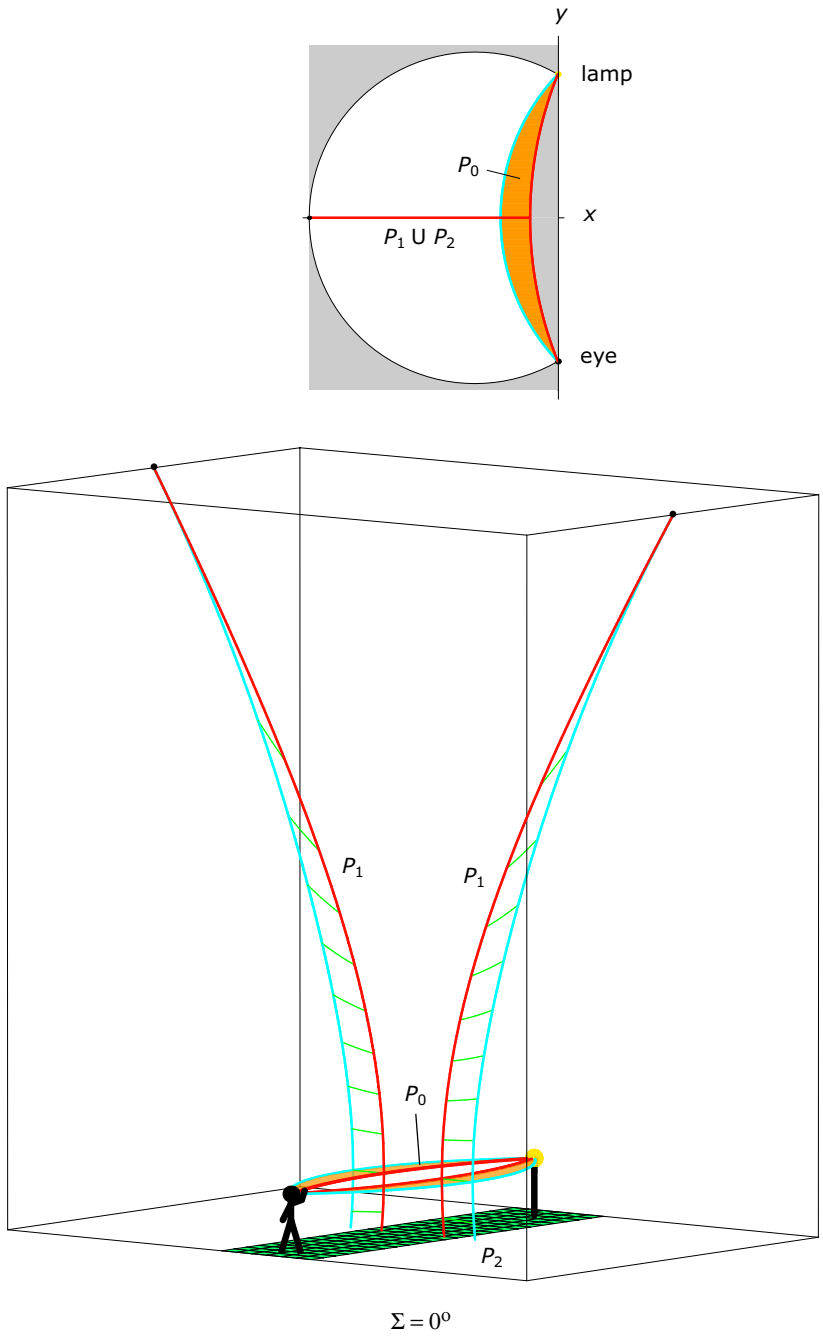

Figure 3.23: Parhelia P_0, superparhelia P_1, and a small piece of the subparhelia P_2, all for $\Sigma = 0$. The super- and subparhelia lie in the xz-plane, equidistant from the eye and the lamp.

Chapter 4

The 22° Halo

The streetlight 22° halo caustic forms in crystals that deviate light through an angle of $D_{\min}(\alpha, n) = 22°$ [Eq. (3.19) with $\alpha = 60°$ and $n' = n = 1.31$]. If we just consider crystals in the vertical plane containing the lamp and the observer, then the halo-making crystals lie on one of two circular arcs as shown in Fig. 4.1. If you revolve either one of these arcs about the observer-lamp line, the surface that you get is the 22° halo caustic, shown in Fig. 4.2. The surface is sometimes known as Minnaert's cigar, after Marcel Minnaert.

> One very cold evening (17°) beautiful halo phenomena could be seen in the steam from a train in a railway-station. Near one of the lamps, where the cloud of steam was blown in every direction, a cigar-shaped surface of light could be seen, having one end near the eye, the other near the lamp; all little crystals traversing this surface were lit up, but the space inside was quite dark...
> M. Minnaert[8, p. 206]

Even in translation, Minnaert's description is charming.

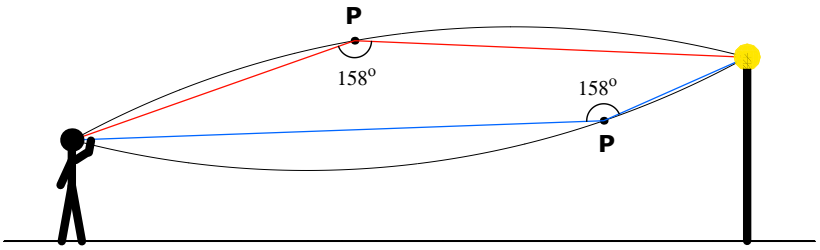

Figure 4.1: Two ray paths (red, blue), both of which are in the plane of the paper and both of which suffer deviations of 22° at points **P**. Any such point **P** lies on one of two circular arcs (black) as shown. The 22° halo caustic is the surface of revolution obtained by revolving either one of these arcs about the observer-lamp line.

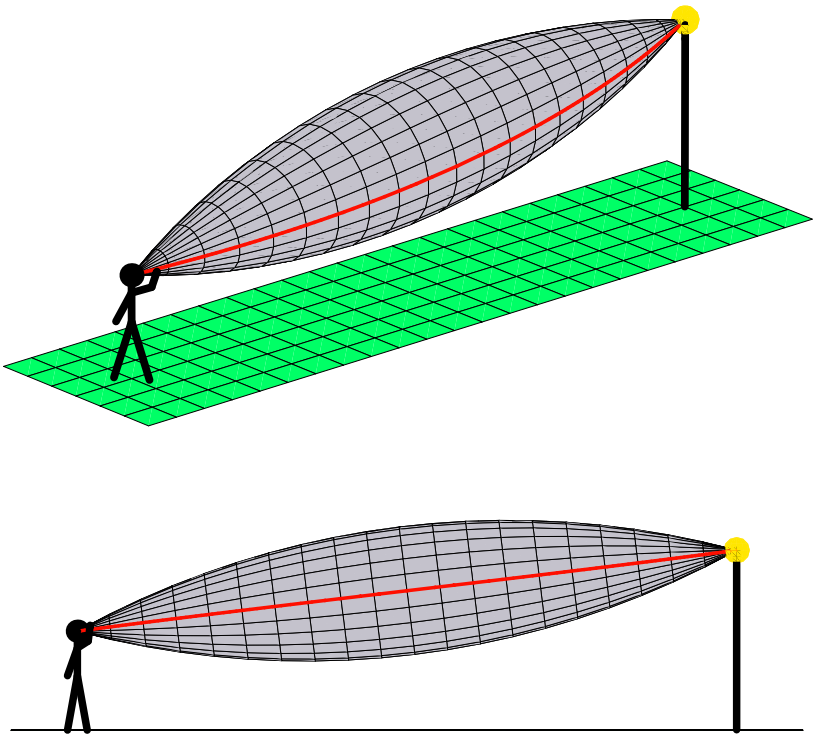

Figure 4.2: The 22° halo caustic surface (gray) together with the right-hand parhelion caustic curve (red).

Chapter 5

Tangent Arcs

The upper tangent arc, or rather its caustic, is shown in Figure 5.1—it is the orange surface on top of the 22° halo caustic. The surface is subtle, since it intersects itself. The self-intersection seems to be responsible for a lot of the interesting behavior of the tanarcs (tangent arcs)—especially the lower tanarc—that occurs when the light source is fairly low.

The tanarcs form in horizontal hexagonal columnar crystals when light rays enter one prism face and exit an alternate prism face. In wedge language: the wedge angle is 60° and the wedge axes are horizontal. The caustic of the tanarcs is distinguished from the tanarcs as a whole by the responsible ray paths: for the caustic the ray paths are symmetric, as in the right-hand diagram of Fig. 3.8. The same is true for the caustics of the 22° circular halo and the parhelia.[1] There are many examples of parhelion caustics in Chapter 3, where the caustics are the minimum deviation curves $d = D_{\min}(a, \Sigma, \alpha, n)$.

5.1 θ-contours

Each blue curve[2] on the tanarc caustic surface in Fig. 5.1 lies in a plane perpendicular to the line through the lamp and the observer's eye. In the side view shown in the lower diagram of the figure, such planes are seen edge-on, and the blue curves therefore appear as line segments. The blue curves do not have much physical significance, but the red curves on the tanarc caustic surface are more interesting:

[1] Also for the super- and subparhelia and for the Wegener arcs. But, in general, halo caustics have nothing to do with symmetric ray paths, so you have to be careful.

[2] These blue curves are the k-contours of Chapter 6.

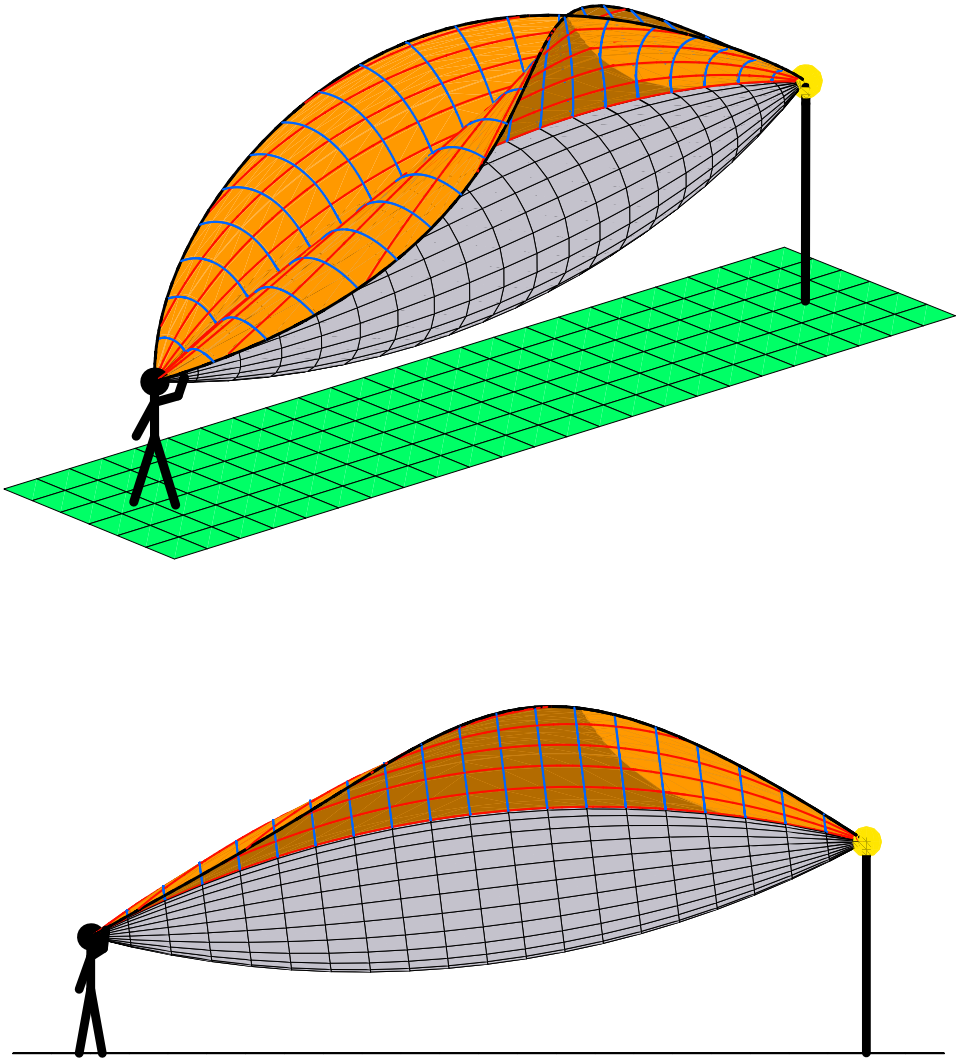

Figure 5.1: Upper tangent arc caustic (orange surface), together with the 22° halo caustic (gray surface), the latter the same as in Fig. 4.2 The lamp elevation with respect to the observer's eye is $\Sigma = 7°$.

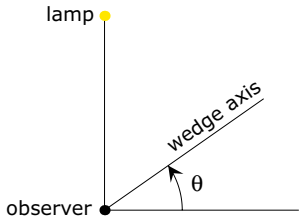

Figure 5.2: Azimuth θ of the (horizontal) wedge axis. The wedge axis is shown passing through the observer, but only the direction of the axis is relevant, not the position. The view is looking down on the lamp and observer.

As we know, the curves and surfaces in spatial halo diagrams are not real physical objects. Rather, they are the locations in space where wedges can send light to the observer's eye in specified ways. Thus the tanarc caustic surface in Fig. 5.1 consists of all points in space where tanarc wedges can refract lamp light to the observer by means of symmetric ray paths.[3] Each red curve shown on the tanarc caustic surface is a contour for some azimuth θ of the wedge axis (Fig. 5.2). So on each θ-contour the axes of the halo making wedges are parallel and point in the direction having azimuth θ. Perhaps you can guess the $\theta = 0°$ contour. It runs along the top of the 22° halo caustic from lamp to eye—it is the trough bottom of the tanarc caustic surface.

The θ-contours in Fig. 5.1 were chosen so that they are more or less equally spaced on the tanarc caustic surface; that choice helps to convey the shape of the surface. The θ-*values* for these contours, however, are far from equally spaced—they are $\theta = 0, \pm 42.6, \pm 56.4, \pm 64.8, \pm 70.8, \pm 75°$. Figure 5.3, on the other hand, has θ-contours whose values are equally spaced. Since in reality all azimuths θ ought to be equally likely, the spacing of the contours on the surface in that figure is a partial indication of halo intensity.

5.2 Trimmed tanarc caustic

Figure 5.4 shows how the tanarc caustic surface in Fig. 5.3 would look to the observer. It takes some study to make the connection between the two figures. Fortunately, much of the complexity in the observer's view arises

[3]The lower tanarc caustic surface should be included. More on this later.

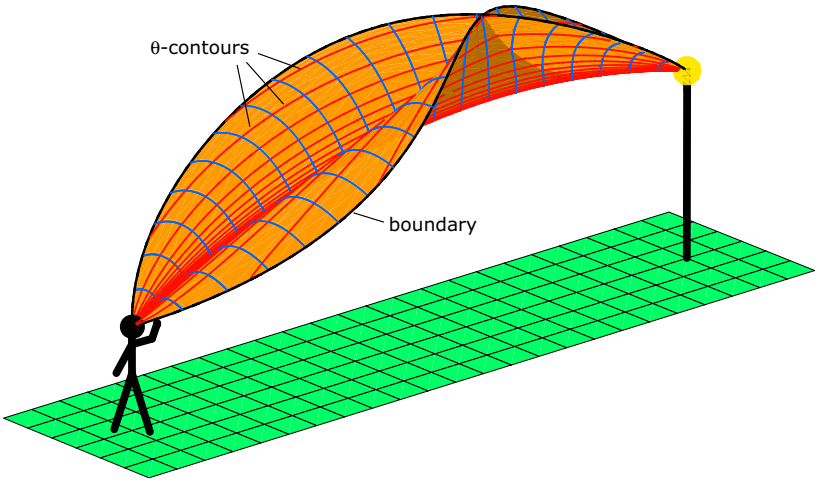

Figure 5.3: The tanarc caustic surface as in Fig. 5.1 but with the θ-contours (red curves) chosen so that their values—rather than the contours themselves—are equally spaced. The spacing of the θ-contours on the surface is therefore a partial indication of halo intensity.

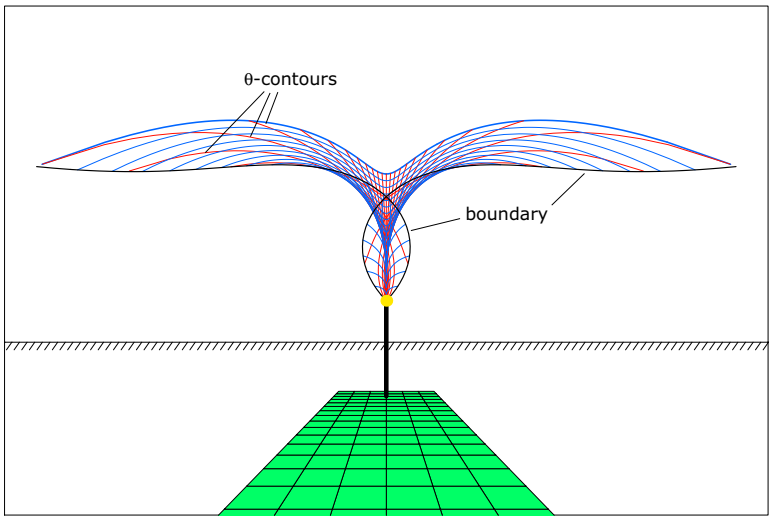

Figure 5.4: The tanarc caustic surface as in Fig. 5.3 but here shown as if photographed by the observer with a 15 mm rectilinear lens. As in Fig. 5.3, the black curve is the boundary of the surface.

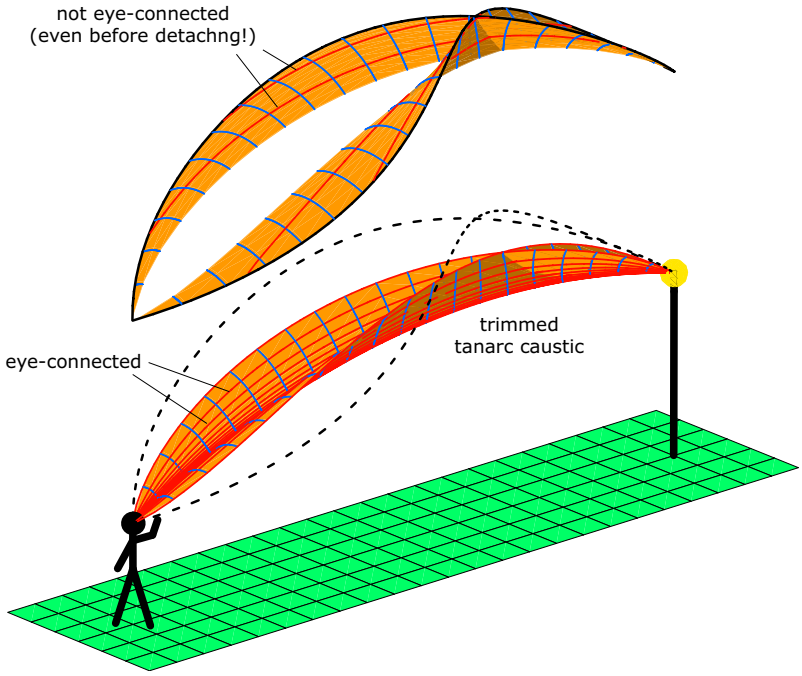

Figure 5.5: Same as Fig. 5.3 but detaching the portion of the tanarc caustic surface whose θ-contours are not eye-connected, that is, that do not reach the observer's eye. The remaining surface is the 'trimmed tanarc caustic.'

in regions that would be too weak to be seen in reality. The tanarc caustic can therefore be trimmed without doing much practical damage.

In Fig. 5.5 a portion of the tanarc caustic surface has therefore been detached, leaving the 'trimmed tanarc caustic.' The detached portion may look like a large chunk to ignore, but it should be of very low intensity. Figure 5.6, the observer's view of the trimmed tanarc caustic, is looking more tractable than that of the untrimmed caustic.

By comparing Fig. 5.5 and Fig. 5.7 you can get some impression of the observer's spatial sense of the tanarc caustic surface. As the observer's eye moves from one blue curve to another in 5.7, starting with the apparently lowest one and ending with the apparently highest, his eye would be moving from the space near the lamp to the space just adjacent to his eye. (But what appears as the highest and longest blue curve in Fig. 5.7 is, in space, the lowest and shortest, as can be seen in Fig. 5.5.)

The trimming of the tanarc caustic that occurred in Fig. 5.5 was de-

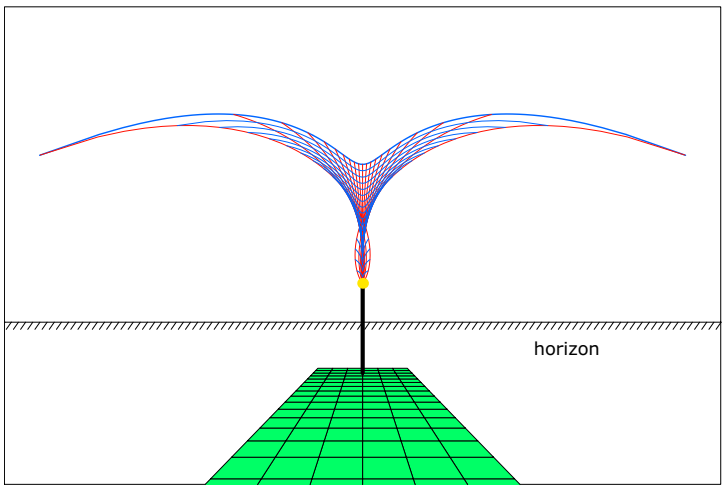

Figure 5.6: The trimmed tanarc caustic of Fig. 5.5 as if photographed by the observer with a 15 mm rectangular lens. This figure is therefore the same as Fig. 5.4 except that it shows only the portion of the tanarc caustic surface whose θ-contours are eye-connected.

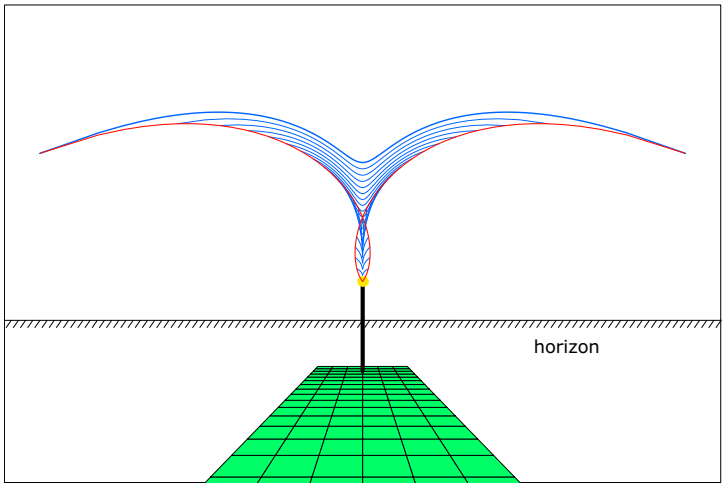

Figure 5.7: Same as Fig. 5.6 but removing all but one of the θ-contours (red), leaving the blue contours. From the apparently lowest blue contour—at the lamp—to the apparently highest, the contours are marching towards us (the observer).

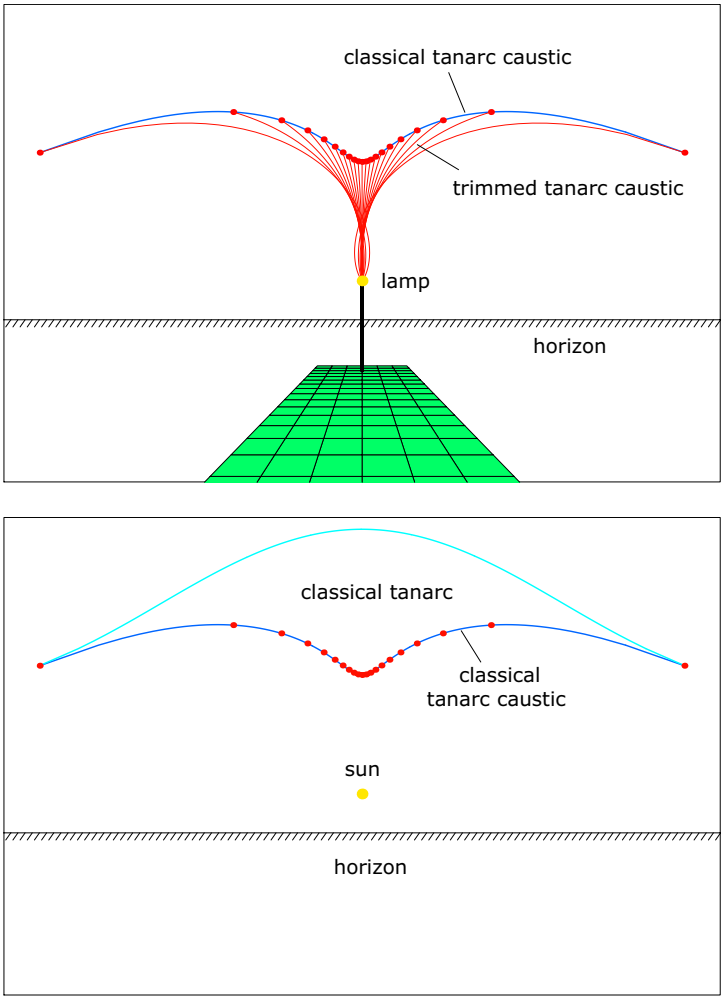

Figure 5.8: Relation of the streetlight tanarc to the classical tanarc. (*Top*) Observer's view of the trimmed tanarc caustic surface of Fig. 5.5. All but one of the blue curves have been removed in order to emphasize the θ-contours (red curves). Because the caustic surface has been trimmed, each θ-contour here is eye-connected. At the apparent end of each of them a red dot has been added. The dots trace out the classical tanarc caustic curve (blue). (*Bottom*) Classical tanarc caustic curve (blue) and classical tanarc for a sun elevation of $\Sigma = 7°$, the same as the lamp elevation in the top diagram.

signed to leave exactly the θ-contours that are eye-connected. Recall from Section 1.2 and Section 2.2 that an eye-connected halo curve, in the observer's view, will appear to end at a point that is the classical instance of the curve.[4] Thus the red dots at the ends of the θ-contours on the tanarc caustic in the upper diagram of Fig. 5.8 should trace out the classical upper tanarc caustic for a sun elevation of 7°, the same as the lamp elevation angle. And indeed they do, as can be seen by comparison with the lower diagram. To the observer, the classical tanarc occupies a region of the celestial sphere on and above the classical tanarc caustic curve, and the streetlight tanarc caustic occupies a region on and below the classical tanarc caustic curve.

A curve is eye-connected exactly when its classical instance exists, that is, when it is non-empty. So the curves that are not eye-connected are just not part of the classical conversation. Those are the curves that got detached in Fig. 5.5.

5.3 Other lamp elevations

As you might expect from the classical tanarc, changing the lamp elevation makes a difference. I want to keep this chapter manageable, so I will look at just a few more lamp elevations, all of them low.[5] Figure 5.9 is like Fig. 5.3 but with $\Sigma = 1°$, so that the lamp is now almost level with the observer's eye. From the reader's perspective the tanarc caustic surface does not look a whole lot different; mainly the self-intersection has moved closer to the observer. The corresponding observer's view of the trimmed tanarc caustic is shown in the lower diagrams of Fig. 5.10.

Figure 5.10 also has the observer's view for $\Sigma = 10°$, and then Fig. 5.11 has observer's views for two more lamp elevations, both negative. Taken together, the diagrams of Fig. 5.10 and Fig. 5.11 show some interesting changes with lamp elevation. The shape of the flare changes somewhat, the distinctive pinching and swelling of the stem changes,[6] and the length of the stem changes, the latter increasing with decreasing lamp elevation. Perhaps we can see some of these features in Aigar Truhin's lovely photo of a Latvian streetlight display taken last December (Fig. 5.13). It will be interesting to

[4] If H is a halo curve or halo surface, then, in the language of Section 2.6, the classical instance of H is H-at-the-eye. The present chapter was written before Section 2.6, so you do not see the 'H-at-the-eye' terminology here.

[5] As of August 2009 there are tanarc caustic diagrams for more lamp elevations in Chapter 6.

[6] The pinching and swelling does not show well in Fig. 5.10 and Fig. 5.11. But see Fig. 5.12.

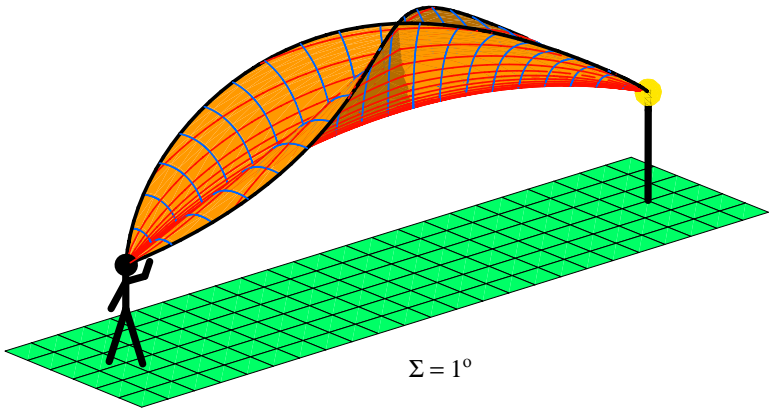

Figure 5.9: Tanarc caustic surface as in Fig. 5.3 but now with lamp elevation $\Sigma = 1°$ instead of $\Sigma = 7°$. The self-intersection in the surface has moved closer to the observer; it would be exactly midway between the observer and the lamp for $\Sigma = 0°$.

try to infer the lamp elevations in his photo from the appearance of the halos, and then to see what Aigar has to say about the actual locations of the lamps.

The prominent white halo at the right in Aigar's photo has a narrow spike going straight up the middle of the flare. On my computer monitor Fig. 5.11 shows the same spike, but the spike disappears when I zoom in. Does the theory predict the spike or not? I don't know. Of course, the spike in the photo could be a well-defined superlamp, but if there is a superlamp mixed in there somwhere, it is not the dominant feature, as the pinching and swelling of the stem in the photo demonstrate.

I don't want to make too much of these diagrams. Please remember they are highly idealized. There is no intensity information in the observer's views other than that which derives from the spacing of the θ-contours. The diagrams assume perfect tilts of the crystals, they ignore the aspect ratios of the crystals, and so forth. I have no idea how, for example, giving the crystals small tilts would affect the stem lengths in the diagrams. So any predictions based on these diagrams should be treated with caution.

For a second example of a real spatial tanarc,[7] consider Jari Luomanen's

[7]This display is described on both the Ice Crystal Halos site [1] of Monday, December 24, 2007, and on Luomanen's site [5].

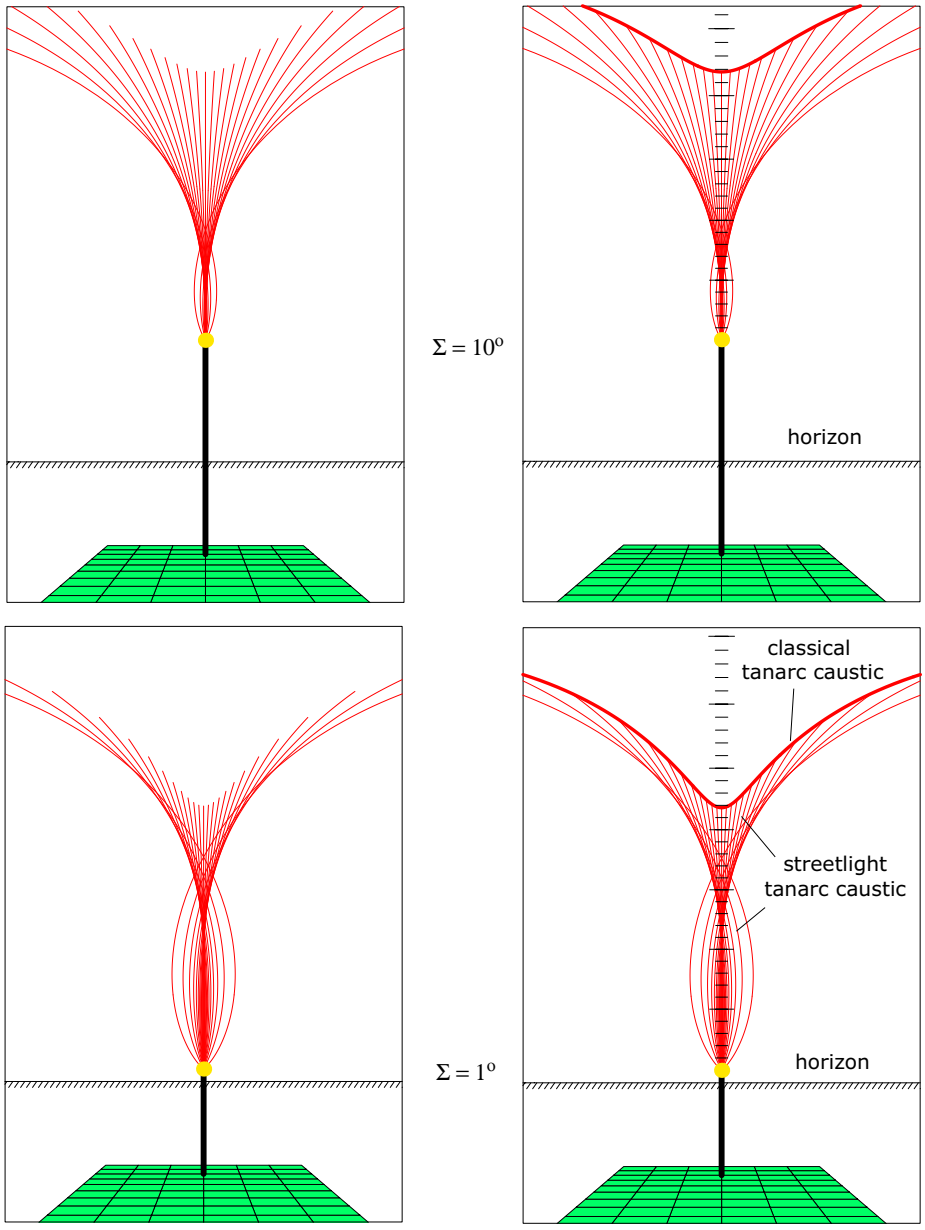

Figure 5.10: Observer's views of trimmed tanarc caustic surfaces for $\Sigma = 10°$ (top) and $\Sigma = 1°$ (bottom). In the right-hand diagrams the classical tanarc caustic curve (thick red) has been added; it is the locus of apparent endpoints of the θ-contours (thin red curves). Tick marks in the right-hand diagrams show that the classical tanarc caustic is where it belongs—21.8° from the lamp. The projection is for a 40 mm rectilinear lens, so the field of view is much smaller than in Fig. 5.6.

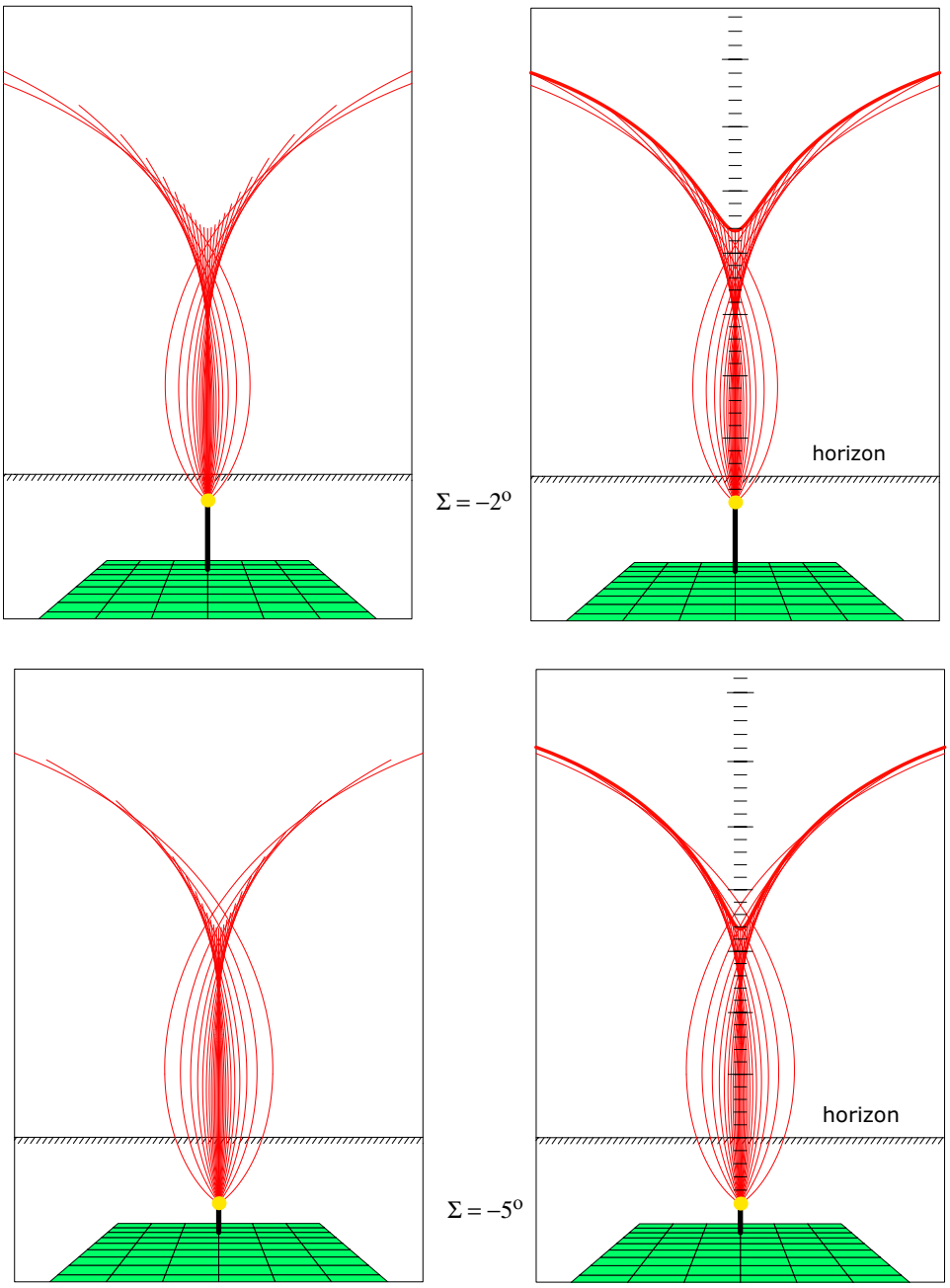

Figure 5.11: Same as Fig. 5.10 but with $\Sigma = -2°$ (top) and $\Sigma = -5°$ (bottom), so that the lamp is somewhat below the observer's eye level. With decreasing lamp elevations the stem lengths have increased from about 6° in the upper diagram in Fig. 5.10 to 17° or 18° in the lower diagram here.

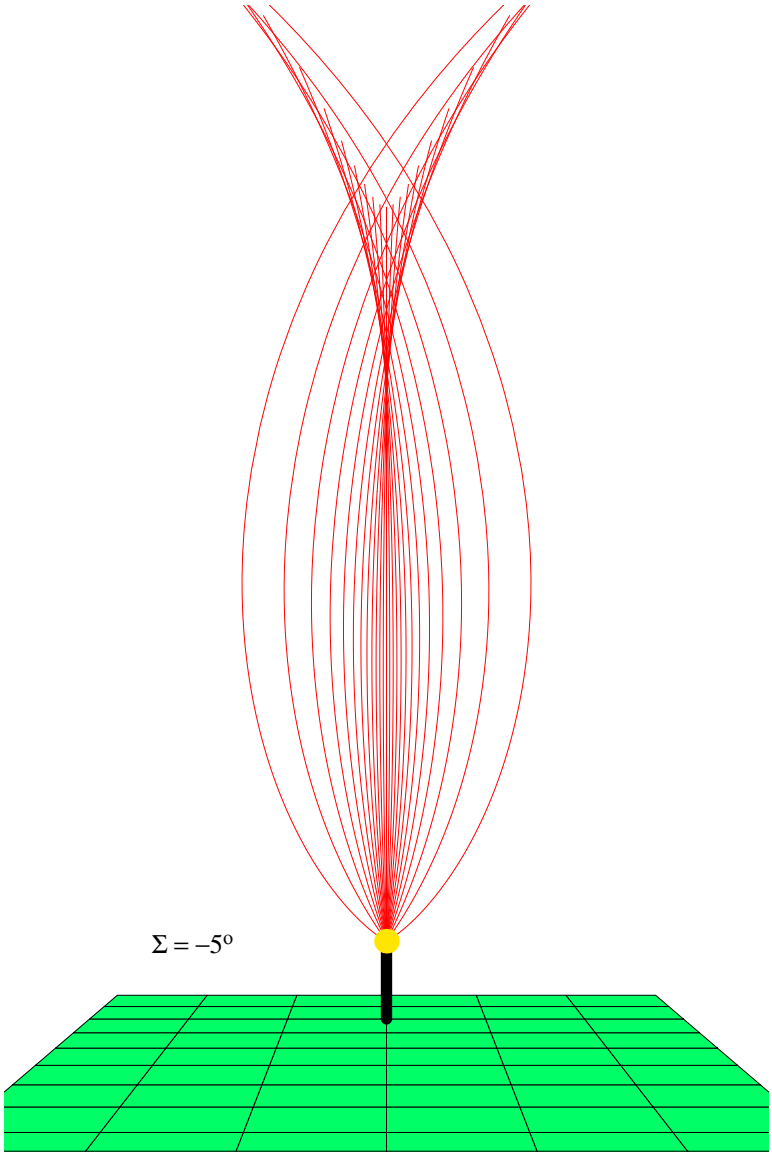

Figure 5.12: Same as the lower left diagram in Fig. 5.11. In this larger version the pinching and swelling of the stem shows better. Step back ten feet from your computer monitor.

Figure 5.13: Streetlight halo display, Sigulda, Latvia, December 28, 2008. ©Aigar Truhin.

photo in Fig. 5.14. Sparkles from individual crystals are only seen[8] in the region of the classical tanarc—the region consisting of the classical tanarc caustic curve and the white area labelled 'interior.' This makes sense, since the sparkles are expected to be coming from crystals close to the camera, exactly where the classical halos are made. Below the classical tanarc caustic curve the halo is homogeneous, with no sparkles, showing that the light is coming from further back towards the lamp. We know that the θ-contours on the tanarc caustic originate on the classical caustic curve, close to the eye, and trail away to the lamp, so we can get some sense of depth even from the flat diagram, without the photo. But we can also get some of that same sense of depth from the photo, thanks to the disposition of the sparkles.

5.4 Halo coloring conventions

The ray path through a halo-making wedge at a point on the tanarc caustic or parhelion caustic is a 'minimum deviation' ray path, that is, it is symmetric with respect to the wedge. That is, the 2-fold symmetry axis of the wedge is also a 2-fold symmetry axis for the ray path. In writing this book, I initially tried to use the color red to indicate minimum deviation, whether loci, ray paths, wedges, etc.

Similarly, I tried to use turquoise to indicate 'maximum deviation.' Maximum deviation occurs when the ray path is tangent to the wedge at entry or exit. You don't see much turquoise in the present chapter, because if a point of the tanarc caustic is a maximum deviation point, then the associated ray path is tangent to the wedge at both entry and exit, by symmetry. In this case I use black rather than turquoise; black is for 'double tangency.' So, for example, the boundary curve in Fig. 5.3 is black, since the halo-making ray paths there are doubly tangential.

I tried to use orange for the interior of the parhelia, and green for the interior of the super- and subparhelia.

The coloring conventions did not always work out. In Fig. 5.1, for example, all the curves and surfaces would have to have been red, if I had followed the red-equals-minimum-deviation convention. And in general I simply ran out of colors if I tried to reserve a special meaning for each color. So I could not achieve the uniform correlation of colors with halos that I had hoped for. There remains enough uniformity, however, especially in Chapters 1–7, that I am mentioning some of the color conventions now.

[8] A few exceptions can be seen near the upper part of the stem. These sparkles are presumably from M-arc crystals. They are intriguing.

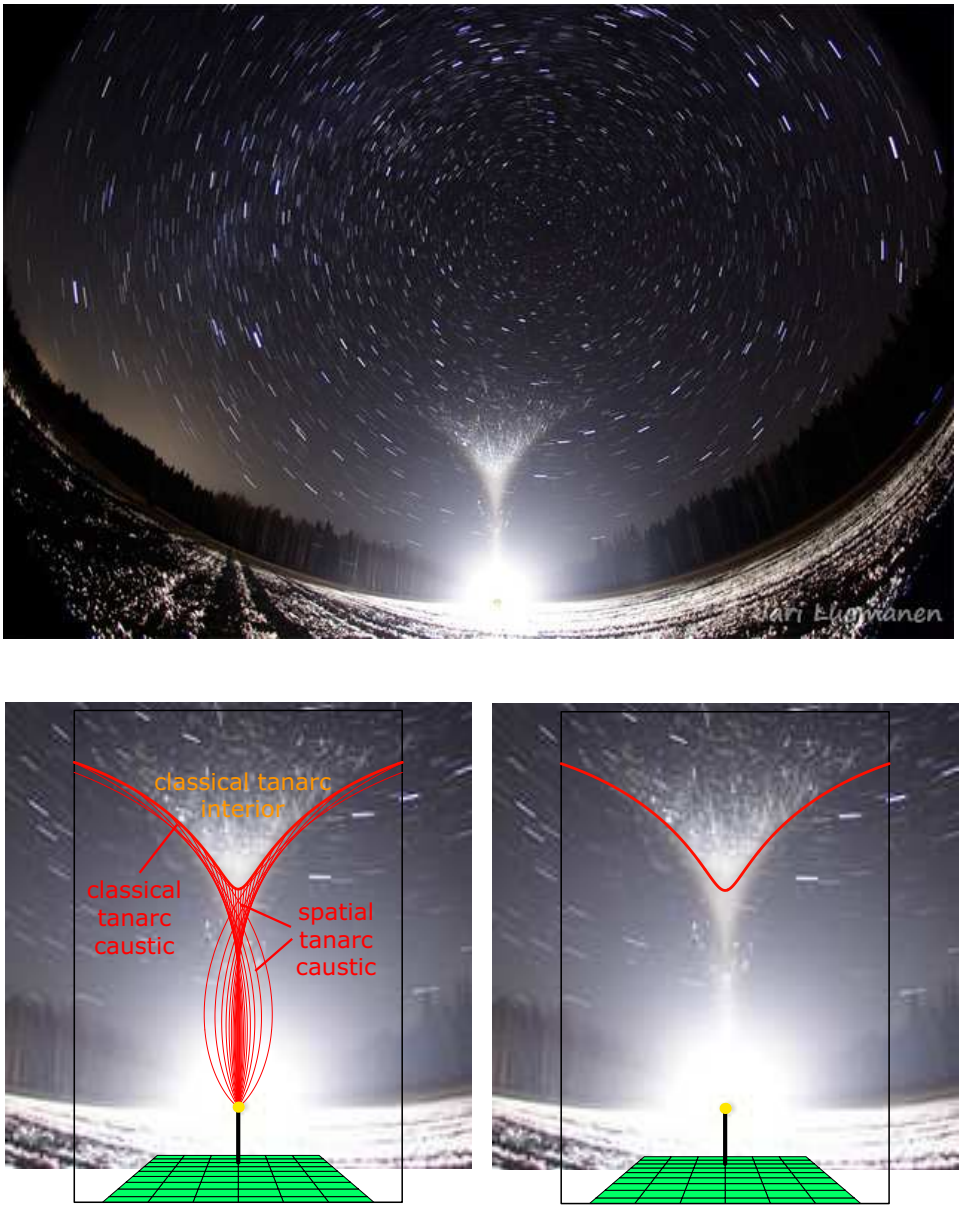

Figure 5.14: (*Top*) Spatial tanarc, Pälkäne, Finland, December 13, 2007. (*Bottom*) Same but with parts of the $\Sigma = -2^\circ$ diagram from Fig. 5.11 superposed. Individual sparkles in the photo are largely confined to the region of the classical tanarc (caustic plus interior), as they should be, since they are coming from crystals close to the camera. All three photos ©Jari Luomanen.

5.5 Tangent arcs from parhelia

Each of the θ-contours in the tanarc caustic surface is a parhelion caustic. For this to be so, 'up' must understood to be in the direction of azimuth θ. You must forget the horizon and forget the direction of the lamp pole, and you must recalculate the effective lamp elevation—it depends on θ. But then the tanarc caustic surface is made up of parhelion caustics for lots of different lamp elevations. If in Fig. 5.3 you study the θ-contours as they progress from the center of the trough outwards, you find the same behavior that was seen for parhelion caustics in Chapter 3 (the $d = D_{\min}(a, \Sigma, \alpha, n)$ curves) for increasing lamp elevations. They move outward, they eventually detach from the lamp and the eye, and they finally reduce to a point and then disappear.

Well, if that wasn't clear, blame Huygens [7, 11]—it was his idea, in the 1660's. Or perhaps refer to the end of Section 7.3, where the construction of the 46° tanarc caustic is described briefly.

5.6 Insights into classical halos

I have learned a lot about the classical halos by thinking about spatial halos. All halos are spatial. Each of us carries around a bunch of exotic curves and surfaces. Some of them are damn big, reaching all the way to the sun. When appropriate crystals pass through the curves and surfaces, they light up.

Here is an example of how, if you think about the more general and supposedly more complex spatial case, you get some insight into the classical case:

In this chapter I talked about the caustic of the upper tanarc, ignoring that of the lower. But the upper and lower tanarcs are really a single halo, and it is artificial to try to separate them. What does the lower tanarc caustic look like—in a streetlight, of course?

In a diagram like Fig. 5.3, you can always interchange the observer and the lamp, leaving all else alone, and you still have a valid diagram; that's just because ray paths are reversible. So go ahead and make the interchange. Now take the resulting diagram and rotate the whole thing through 180° about the horizontal line that is a perpendicular bisector of the line segment between the observer and the lamp. The observer and lamp are now both back in their original positions. If you performed that same rotation on all the crystals and ray paths, then you still have a valid diagram. Moreover, horizontal crystals went to horizontal crystals, and tanarc caustic ray paths

went to tanarc caustic ray paths. In short, you end up with the lower tanarc caustic surface. For each lamp elevation, the caustic of the lower tanarc differs from that of the upper only by a 180° rotation. You'd never guess, thinking classically.

For lamp elevations that are fairly low but positive, the self-intersection that we saw in the upper tanarc caustic surface in Fig. 5.3 is, as in the figure, up near the lamp end of the caustic surface. Classically, the observer can't see the lamp end of the surface; he sees only the part that is right up against his eye. For those same lamp elevations, though, the self-intersection in the lower tanarc caustic surface is down at the observer's end of the surface, making its presence known even in the classical instance, giving rise to those loops and cusps that always seemed so peculiar.

Chapter 6

From Tanarcs to Circumscribed Halo

This chapter is a continuation of Chapter 5, where the upper tanarc caustic surface was shown for several lamp elevations, all 10° or less. Figures 6.2–6.6 show both the upper and lower tanarc caustics for lamp elevations ranging from 15° to 32°. The figures illustrate the evolution of the tanarc caustics from two separate surfaces to a single merged surface, the latter usually known as the circumscribed halo. Figures 6.7–6.11 illustrate the same evolution, but for the observer's view.

Unlike many of the figures in Chapter 5, the tanarc caustic surfaces shown here are complete, not trimmed. That makes most of the surfaces complicated, especially as seen by the observer.

I realize that the figures are not very successful in conveying the shape of the surfaces. It helps to recall that very near the eye the surface should resemble its classical instance. So in Fig. 6.2, for example, the lower tanarc caustic surface near the eye should have the distinctive looped shape of the classical lower tanarc caustic curve shown in Fig. 6.7. And then by symmetry (Section 5.6) the upper tanarc caustic surface near the lamp should have the same looped shape, but rotated 180°. To help with the perspective in the observer's views, remember that the classical instance of the halo is always closest to the observer; from there, each eye-connected θ-contour runs back toward the lamp.

In addition to the contours of θ (red) in the figures, there are contours of k (blue). The k-coordinate of a point in space is the scalar projection of the point onto the observer-lamp line. Level surfaces of k are therefore planes perpendicular to the observer-lamp line. As far as I know, the

k-contours have no physical significance, but they help with the perspective in the observer's views, if you compare the observer's view with the 'reader's view'—any view with a generic perspective.

As usual, there is only limited intensity information in the figures. The halo displays should be weak near the boundary curves, since the halo-making ray paths on the boundary are doubly tangential. And the displays should be weak where, in the observer's view, contours are widely spaced.

6.1 Merging

Classically, the upper and lower tanarc caustics merge at a sun elevation of $29.25°$ to form the circumscribed halo caustic. The value $29.25°$ is the lowest lamp elevation for which the $\theta = 90°$ contour is eye-connected. For lamp elevations $\Sigma \geq 29.25°$ all of the θ-contours are eye-connected; compare, say, Fig. 6.4, where $\Sigma = 22°$, with Fig. 6.6, where $\Sigma = 32°$. Also compare the corresponding observer's views, Fig. 6.9 and Fig. 6.11, and remember that the θ-contours that seem to end on the classical tanarc caustic curve are eye-connected, whereas those that end on the boundary curve are not.

The streetlight tanarc caustics merge at the much lower lamp elevation of $15.64°$; this is the lowest lamp elevation for which the $\theta = 90°$ contour is non-empty. But at this lamp elevation the contour is not yet eye-connected, so classically it is invisible. Figure 6.2 and Fig. 6.3, and Fig. 6.7 and Fig. 6.8, illustrate the transition from separate to merged.

The figures in this chapter all have θ-contours with $\theta = 0, \pm 10, \pm 20, \pm 30, \ldots$. Some of the θ-contours are labeled, so that, by following them from one figure to the next, you can see how the merging of the upper and lower tanarc caustics comes about.

The lamp elevations $15.64°$ and $29.25°$ for merging are highly theoretical, and in reality you will not see the tanarcs merge at lamp elevation $15.64°$, because the arcs are so weak where they merge. But I wonder whether we will be able to see the merging at some lamp elevation greater than $15.64°$ but still less than the classical merge value of $29.25°$, as for instance the lamp elevation $\Sigma = 27°$ of Fig. 6.5. It will not happen, however, if θ-contours that are not eye-connected are really as weak as I claimed in Chapter 5.

6.2 Some tanarc mathematics

As explained briefly in Section 5.5, the tanarc caustic surface is made up of parhelion caustic curves, each with a funny 'up' direction. That's because if

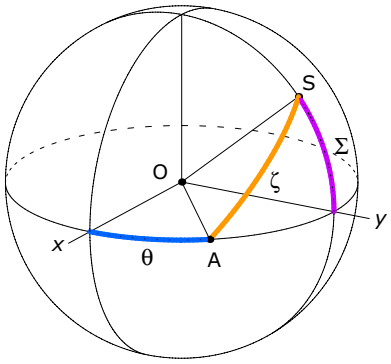

Figure 6.1: Celestial sphere with sun (or lamp direction) **S** and with point **A** on the equator. The angle Σ (purple) is the sun elevation, and the angle θ (blue) is the azimuth of **A**. The angle ζ (orange) is the effective zenith angle of **S** when **A** is considered to be the zenith. Hexagonal prismatic crystals having principal axis **A** make 'parhelia' whose effective sun elevation (not shown) is $\Sigma_\theta = 90 - \zeta$.

you rotate a horizontal hexagonal columnar crystal about its (principal) axis, you get a pair of 'parhelia' whose effective zenith direction is not the ordinary zenith but rather the direction of the crystal axis. The two parhelion caustic curves make up a θ-contour of the tanarc surface.

On the celestial sphere in Fig. 6.1, the point **S** is the sun (or lamp), and the point **A** is in the direction of the (horizontal) crystal axis. When **A** is considered as the zenith, the effective zenith angle of the sun is $\zeta = \angle\mathbf{AOS}$. If θ, $-90° \le \theta \le 90°$, is the azimuth of the crystal axis as usual, and if $P(\theta)$ ('P' for 'parhelia') is the corresponding θ-contour, then $P(\theta)$ is indeed just a pair of parhelia. We therefore know from Section 3.3 that the nature of $P(\theta)$ depends on the effective sun elevation $\Sigma_\theta = 90 - \zeta$:

$$P(\theta) \text{ is eye-connected} \quad \text{iff} \quad |\Sigma_\theta| \le \Sigma_1 \tag{6.1}$$

$$P(\theta) \text{ is non-empty} \quad \text{iff} \quad |\Sigma_\theta| \le \Sigma_2, \tag{6.2}$$

where Σ_1 and Σ_2 are given by Eq. (3.41) and Eq. (3.42).

From Fig. 6.1

$$\mathbf{S} = (0,\, \cos\Sigma,\, \sin\Sigma) \tag{6.3}$$

$$\mathbf{A} = (\cos\theta,\, \sin\theta,\, 0), \tag{6.4}$$

so

$$\sin\Sigma_\theta = \cos\Sigma\, \sin\theta. \tag{6.5}$$

Thus if $\theta_1(\Sigma, \alpha, n)$ and $\theta_2(\Sigma, \alpha, n)$ are the θ-values corresponding[1] to Σ_1 and Σ_2, respectively, that is, if, for $i = 1, 2$,

$$\sin \Sigma_i = \cos \Sigma \sin \theta_i \quad \text{if} \quad \Sigma \leq 90 - \Sigma_i \tag{6.6}$$

$$\theta_i = 90^\circ \quad \text{if} \quad \Sigma \geq 90 - \Sigma_i, \tag{6.7}$$

then

$$P(\theta) \text{ is eye-connected} \quad \text{iff} \quad |\theta| \leq \theta_1 \tag{6.8}$$

$$P(\theta) \text{ is non-empty} \quad \text{iff} \quad |\theta| \leq \theta_2 \tag{6.9}$$

If $\Sigma = 15^\circ$, for example, then Eq. (6.6) gives $\theta_1 = 64.6^\circ$ and $\theta_2 = 85.6^\circ$. The $\theta = 70^\circ$ and $\theta = 80^\circ$ contours in Fig. 6.2 should therefore be detached from the eye (they are), and the $\theta = 90^\circ$ contour should be missing (it is).

Figure 5.5 in Chapter 5 showed the trimmed portion of the $\Sigma = 7^\circ$ upper tanarc caustic surface as well as the portion that was detached. You can see now how the figure was made. The trimmed (i.e., remaining) portion is the portion with $|\theta| \leq \theta_1$, and the detached portion is the portion with $\theta_1 \leq |\theta| \leq \theta_2$. For $\Sigma = 7^\circ$ you find $\theta_1 = 61.5^\circ$ and $\theta_2 = 76^\circ$.

Merging revisited

The streetlight upper and lower tanarc caustics overlap (i.e, have non-empty intersection) when the $\theta = 90^\circ$ contour is non-empty. According to Eq. (6.2), that happens when $\Sigma_{90^\circ} \leq \Sigma_2$, i.e., $90 - \Sigma \leq \Sigma_2$, i.e., $\Sigma \geq 90 - \Sigma_2$.

Similarly, the classical upper and lower tanarc caustics overlap when the $\theta = 90^\circ$ contour is eye-connected, and that happens when $\Sigma \geq 90 - \Sigma_1$. To summarize:

The classical upper and lower tanarc caustics merge when $\Sigma = 90 - \Sigma_1$.
The streetlight upper and lower tanarc caustics merge when $\Sigma = 90 - \Sigma_2$.

For the usual case where $\alpha = 60^\circ$ and $n = 1.31$, Eq. (3.41) and Eq. (3.42) give $90 - \Sigma_1 = 29.25^\circ$ and $90 - \Sigma_2 = 15.64^\circ$. Those are the lamp elevations for merging that were discussed in Section 6.1.

[1] This characterization of θ_i as corresponding to Σ_i is only correct for the case $\Sigma \leq 90 - \Sigma_i$. If $\Sigma > 90 - \Sigma_i$, then there is no θ (axis azimuth) corresponding to Σ_i.

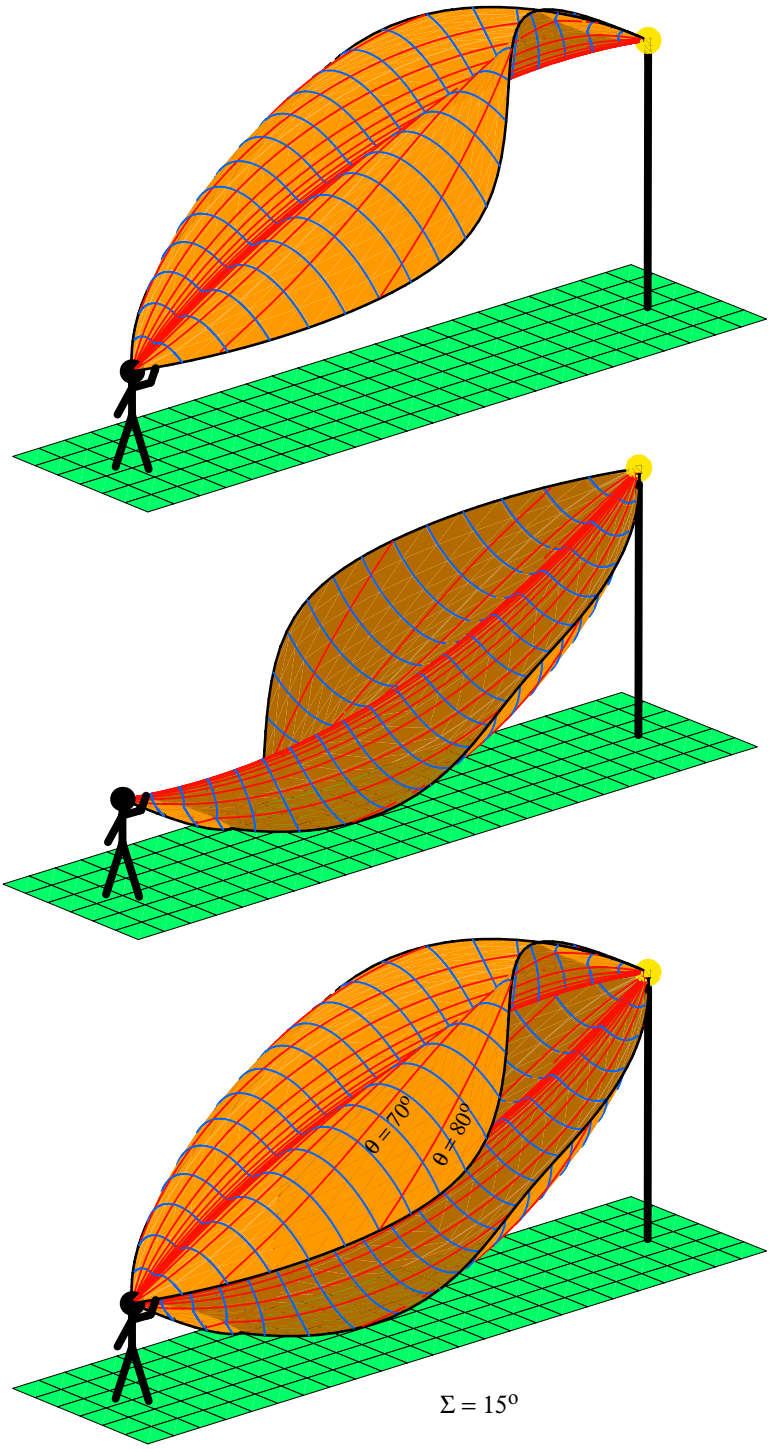

Figure 6.2: (*Top*) Upper tanarc caustic for a lamp elevation of $\Sigma = 15°$. (*Middle*) Lower tanarc caustic. (*Bottom*) Both together. The interval between θ-contours (red) is $\Delta\theta = 10°$.

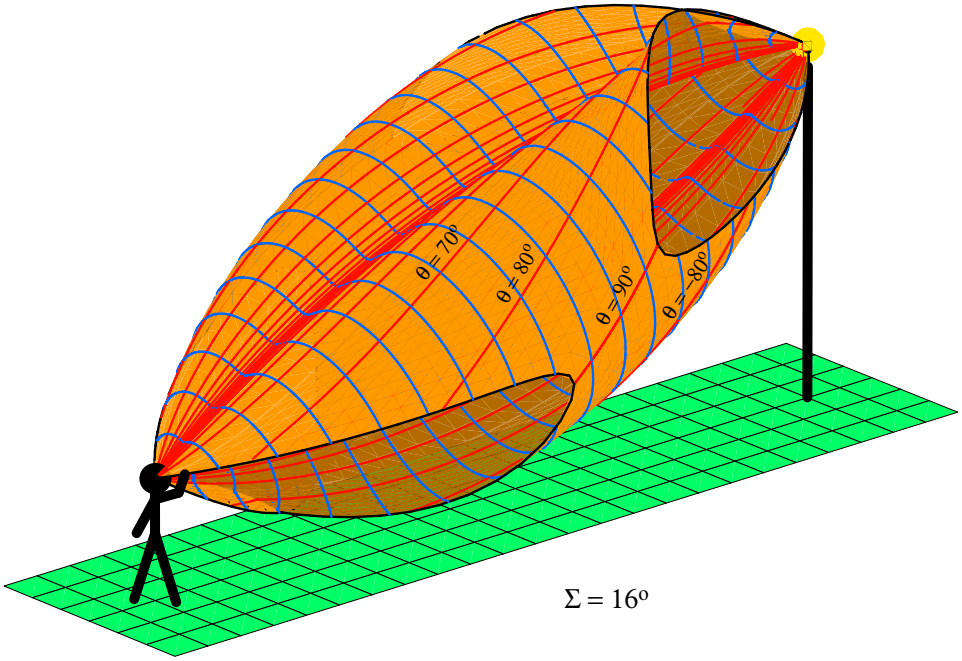

Figure 6.3: Upper and lower tanarc caustic surfaces for $\Sigma = 16°$. In contrast to the previous diagram, the $\theta = 90°$ contour here is non-empty, and so the two surfaces have merged.

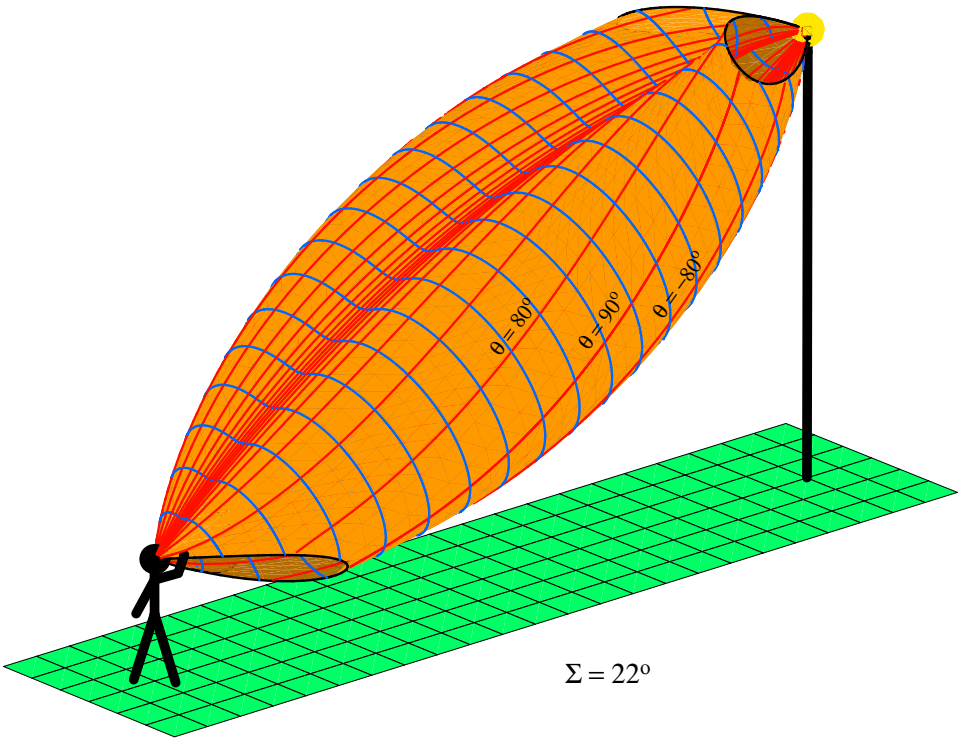

Figure 6.4: Upper and lower tanarc caustics for $\Sigma = 22°$.

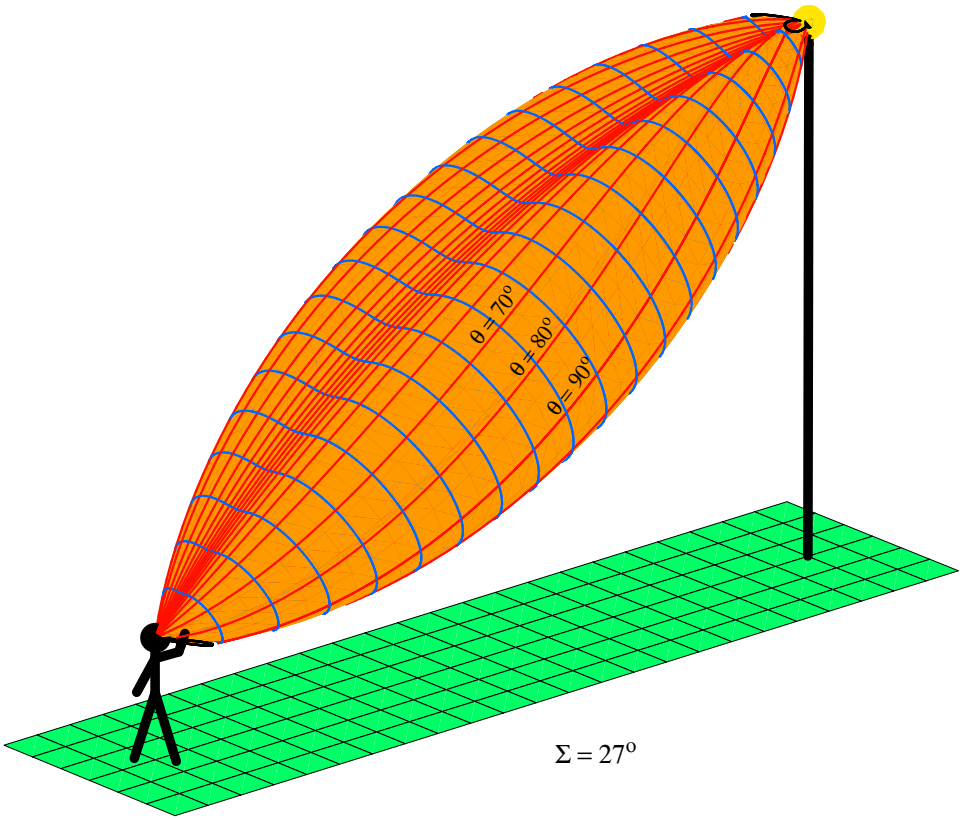

Figure 6.5: Upper and lower tanarc caustics for $\Sigma = 27°$. The black boundary curves that were evident in Figs. 6.2–6.4 have now shrunk dramatically but are still visible—at the eye and at the lamp.

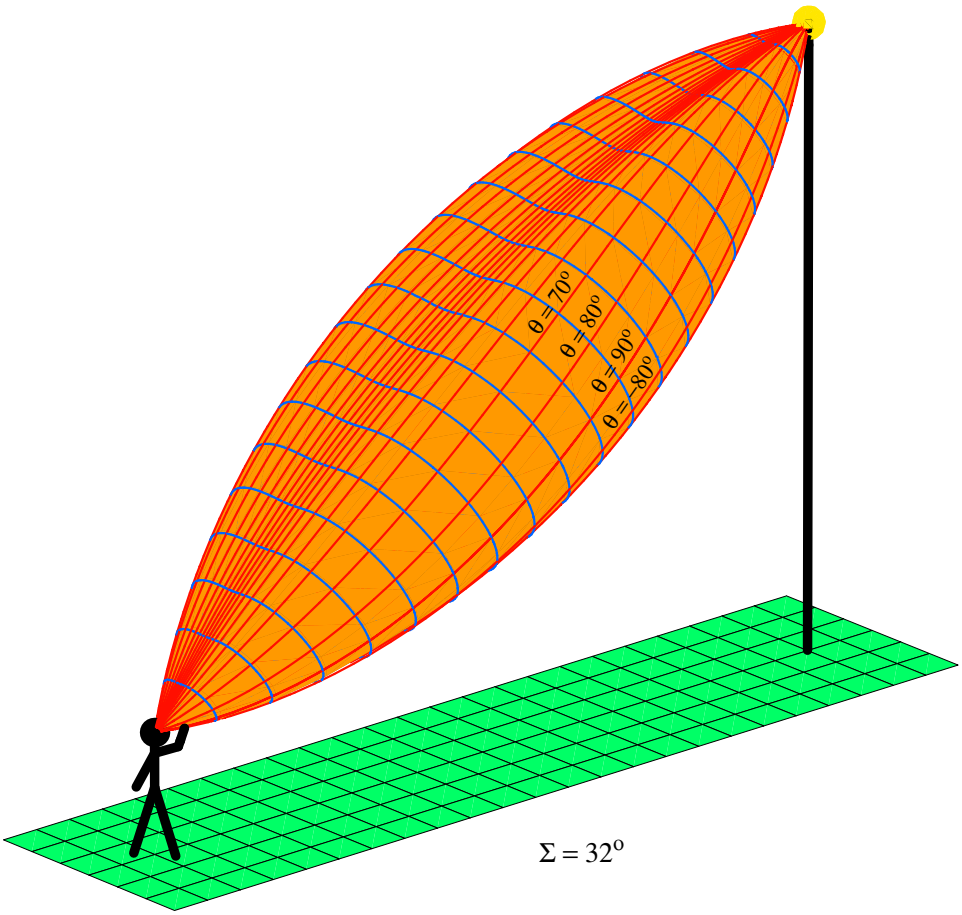

Figure 6.6: Upper and lower tanarc caustics (or circumscribed halo caustic) for $\Sigma = 32°$. In contrast to Figs. 6.2–6.5, all θ-contours here are eye-connected.

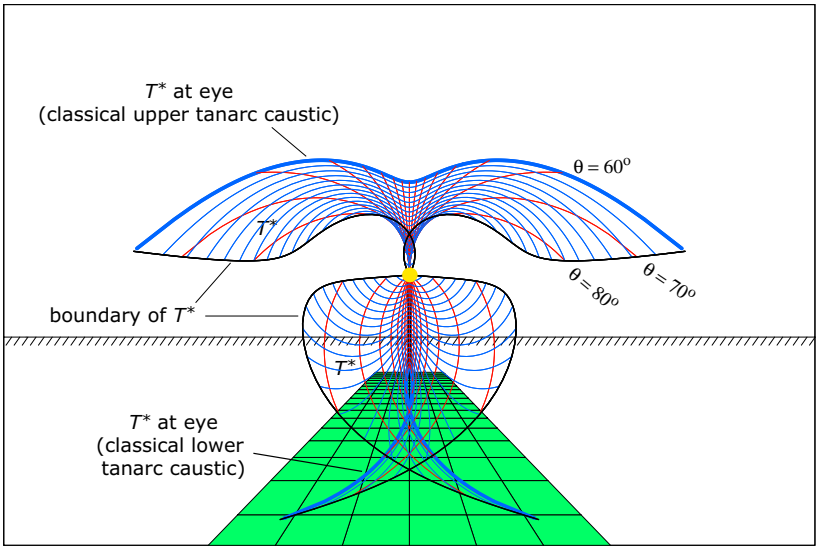

Figure 6.7: The $\Sigma = 15°$ tanarc caustic surface T^* as if photographed by the observer with a 10 mm rectilinear lens. The θ-contours (red), the k-contours (blue), and the boundary curves (black) are all the same as in Fig. 6.2.

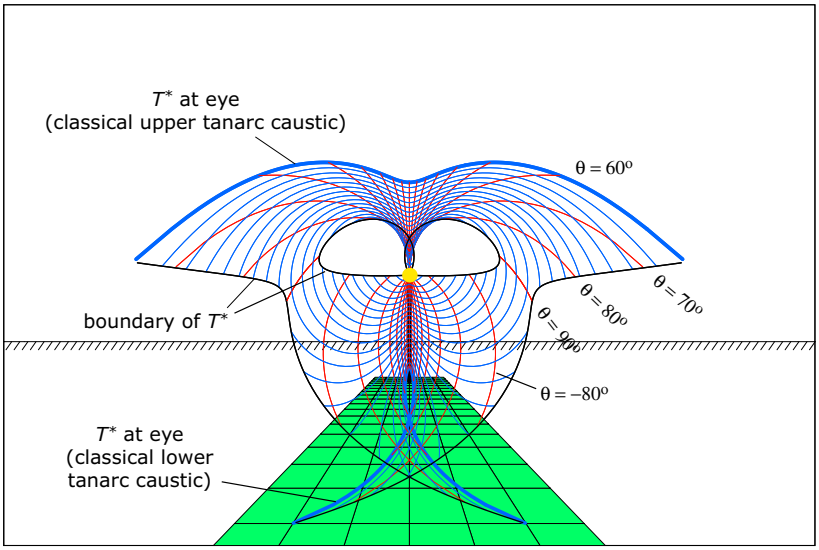

Figure 6.8: The $\Sigma = 16°$ tanarc caustic surface T^* as if photographed by the observer with a 10 mm rectilinear lens. The streetlight upper and lower tanarc caustics have merged, but not their classical instances. Figure 6.3 is the corresponding 'reader's view.'

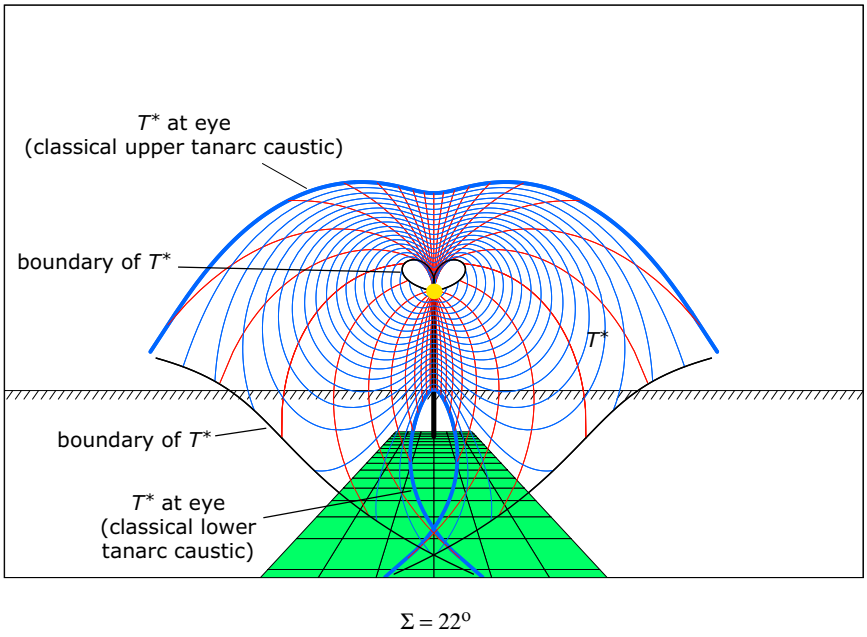

Figure 6.9: The $\Sigma = 22^\circ$ tanarc caustic surface T^* as if photographed by the observer with a 10 mm rectilinear lens. Figure 6.4 is the corresponding reader's view.

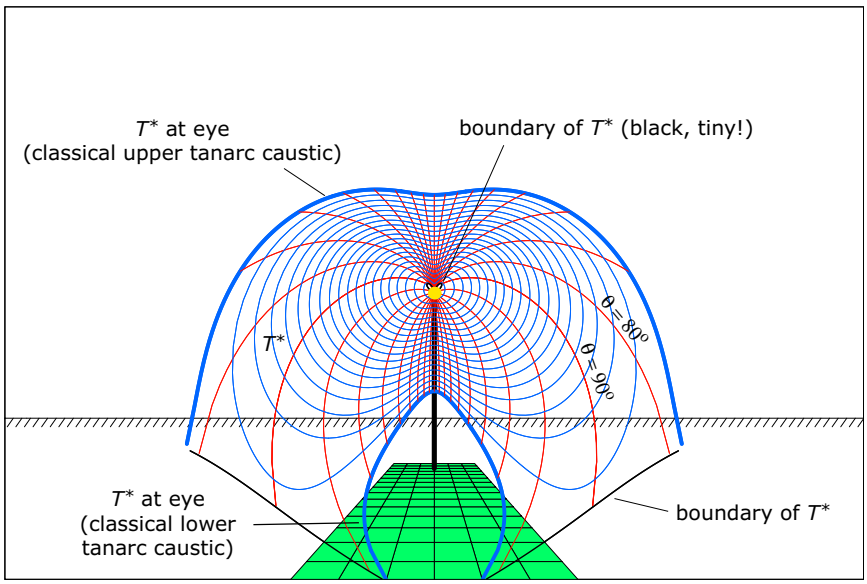

Figure 6.10: The $\Sigma = 27°$ tanarc caustic surface T^* as if photographed by the observer with a 10 mm rectilinear lens. Although the boundary curves (black) are tiny in the reader's view (Fig. 6.5), those near the eye look big to the observer. The boundary curves at the lamp are also visible here, barely, if you zoom in.

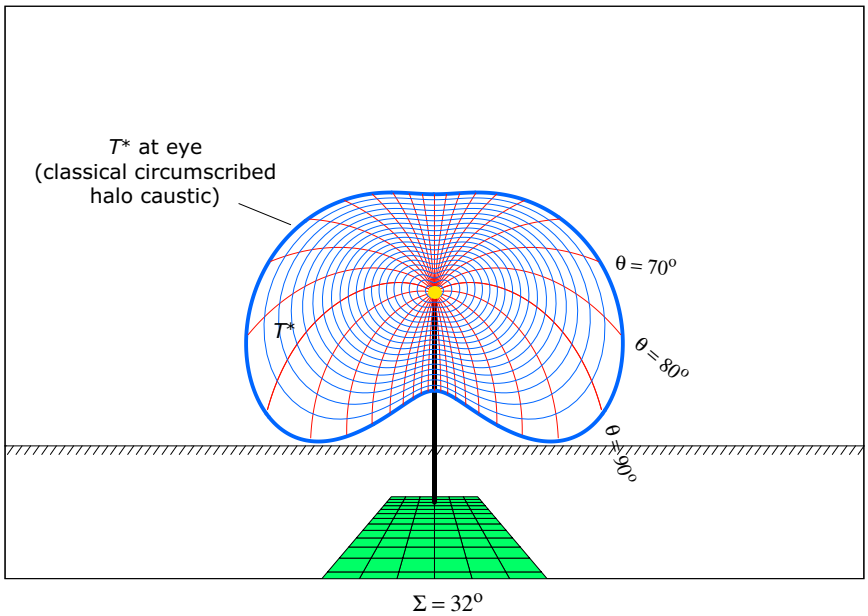

Figure 6.11: The $\Sigma = 32°$ circumscribed halo caustic surface T^* as if photographed by the observer with a 10 mm rectilinear lens. The classical tanarc caustic curves have finally merged. The corresponding reader's view is Fig. 6.6.

Chapter 7

Circumzenith Arc

The circumzenith arc is formed in horizontal hexagonal plate crystals by light rays that enter the top basal face of a crystal and exit a prism face. In wedge language: wedges for the circumzenith arc have wedge angle 90° and have their entry faces horizontal. Figure 7.1 shows a circumzenith arc formed in a streetlight. It is the surface in space consisting of all points where wedges can send lamp light to the observer as just described.

The circumzenith arc in Fig. 7.1 gives the appearance of being truncated, but it is all there. Its boundary consists of the purple and blue curves. At each point of the purple curve the ray path through the halo-making wedge is tangent to the wedge at entry; the entry ray is thus horizontal, and the purple curve is at the same level as the lamp. At each point of the blue curve it is the exit ray that is tangent to the wedge; the ray is vertical when projected to the normal plane of the wedge. Ray paths that are nearly tangential, whether at entry or exit, are weak, and the circumzenith arc near its boundaries is not apt to be perceptible. Do not be unduly distracted by the boundary curves.

At each point of the circumzenith arc between the red curve and the purple curve, the ray path through the halo-making wedge is more nearly tangential at entry than at exit. Between the red curve and the blue curve, the ray paths are more nearly tangential at exit than at entry. On the red curve itself the ray paths are symmetric.

Each green curve on the surface is a θ-contour. As explained in Section 5.1, on each θ-contour the axes of the halo-making wedges are parallel and have azimuth θ. Of course now the halo in question is the circumzenith arc rather than the tanarc caustic.

Figure 7.2 shows the circumzenith arc of Fig. 7.1 as seen by the observer.

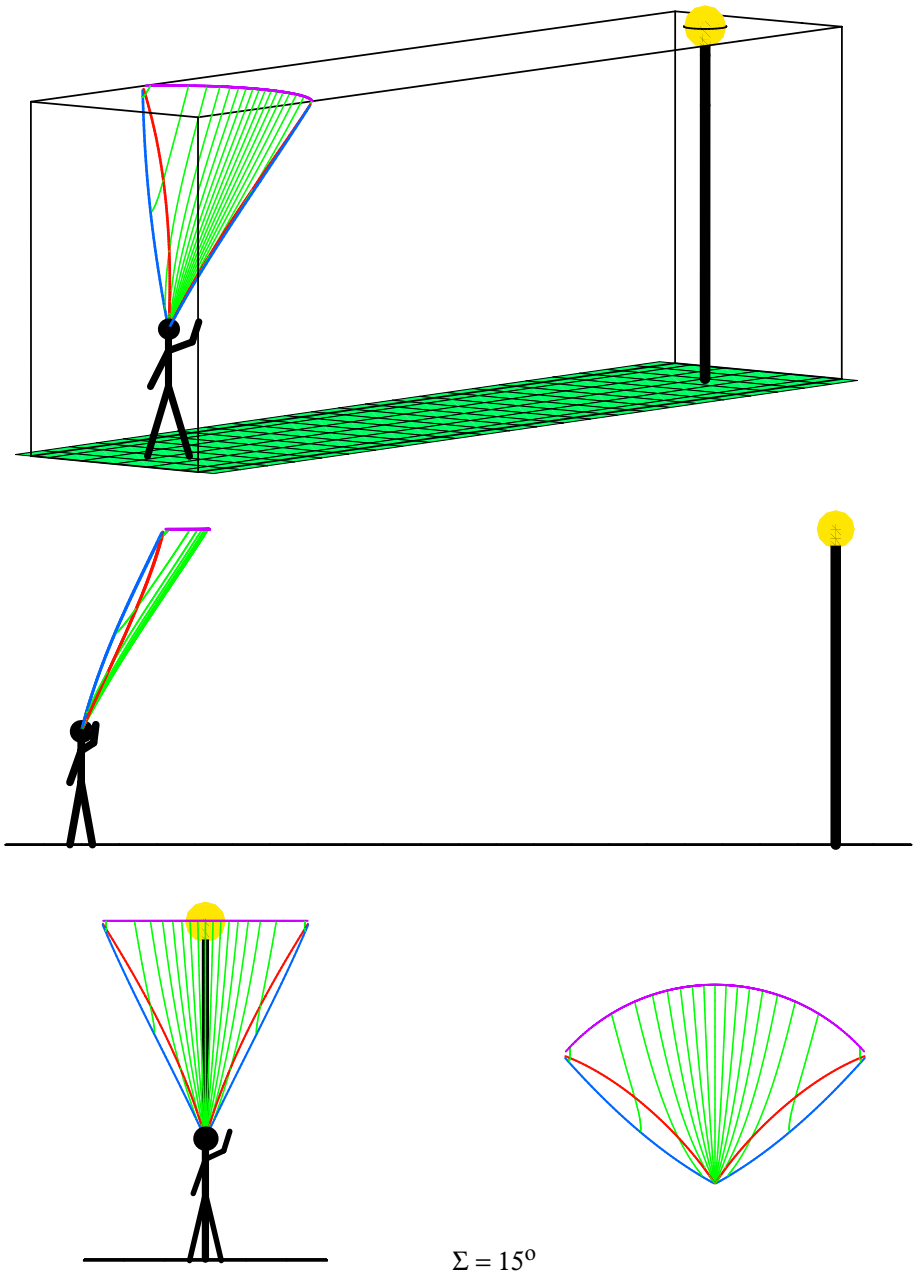

Figure 7.1: Streetlight circumzenith arc. In spite of the terminology, it is a surface, not an arc. The entire surface is shown; it has not been truncated. In the lower left diagram the view is from directly behind the observer and looking toward the lamp. The lower right diagram is a top view. The lamp elevation is $\Sigma = 15°$.

My brain requires some forcing for it to accept that the purple boundary curve here is actually the highest part of the surface. It is in fact at the level of the lamp; from there the surface bends down toward us. The green dots can be thought of, loosely, as plastered up against the eye.

As with all spatial halos, the eye-connected curves are subtle but important. Each eye-connected curve appears to the observer to end at a point in space which is in the direction of the classical halo point of the curve. Each green dot in Fig. 7.2 is the classical halo point for an eye-connected green curve. The green dots thus trace out the classical circumzenith arc, the familiar circumzenith arc that would form in sunlight. Tick marks show that it is located as expected, about 47° from the lamp.

7.1 Twins

I have seen the streetlight circumzenith arc many times, but I rarely remember to look for the circumhorizon arc. The classical circumhorizon and circumzenith arcs occur at high and low sun elevations, respectively, and hence are not seen together. But the streetlight versions occur simultaneously, at the same lamp elevation, and the two surfaces are even congruent, as illustrated in Fig. 7.3. In fact, they differ from each other only by a 180° rotation. The axis of rotation is the horizontal line that perpendicularly bisects the line segment between the eye and the lamp.

To see why the two halos only differ by this 180° rotation, start with a crystal and ray path that are contributing to the circumzenith arc. First interchange the lamp and the eye, without tampering with the crystal or ray path. Since light ray paths are reversible, you still have a bona fide ray path. The new entry face, however, is the old exit face, namely, a vertical prism face, and the new exit face is the old entry face, namely, the top horizontal basal face. Now perform the 180° rotation on the entire configuration—crystal, ray path, eye, and lamp—thus returning the eye and lamp to their original positions. The crystal still has its basal faces horizontal, though interchanged. The new ray path is indeed a ray path from lamp to eye. The entry face is a vertical prism face and the exit face is the bottom horizontal basal face; the ray path is a ray path for the circumhorizon arc.

Although the circumhorizon and circumzenith arcs are congruent, they will not look the same to the observer. I can't remember whether I have seen them simultaneously in a real display, and I look forward to watching for them more carefully next winter.

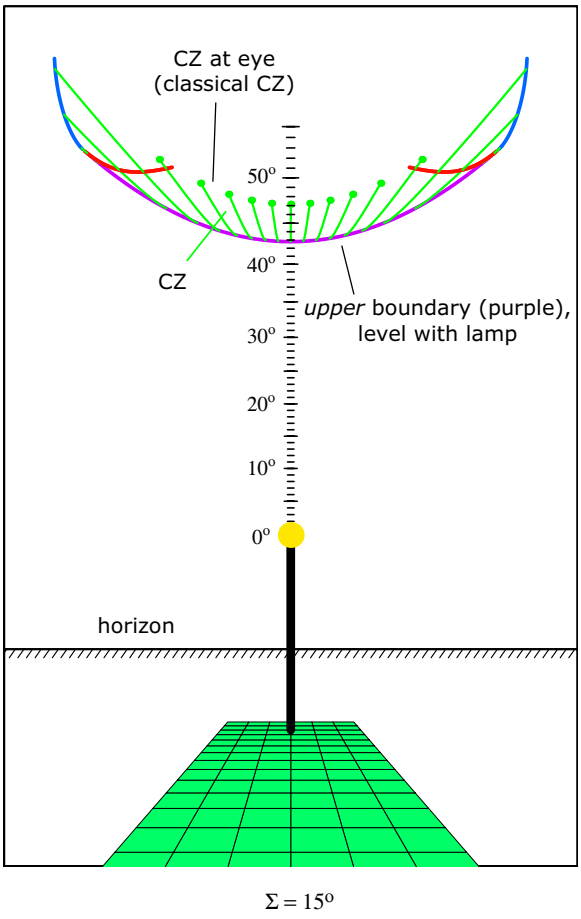

Figure 7.2: The circumzenith arc CZ of Fig. 7.1 as if photographed by the observer with a 15 mm rectilinear lens. The green dots, one for each of the green curves that is eye-connected, are the classical halo points—they trace out the classical circumzenith arc, which, in the language of Section 2.6, is the circumzenith-arc-at-the-eye.

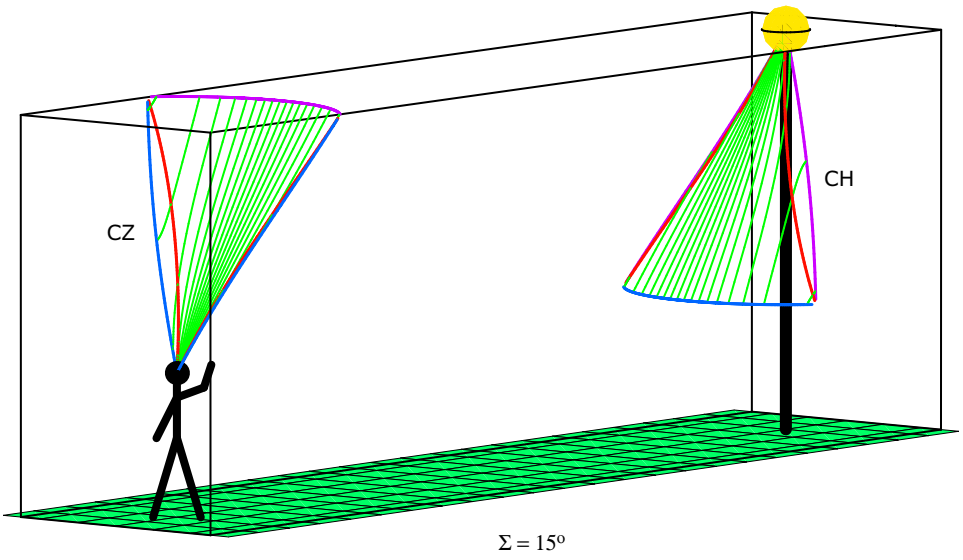

Figure 7.3: Same as Fig. 7.1 but with the addition of the circumhorizon arc CH. It differs from the circumzenith arc CZ by a 180° rotation.

7.2 Dependence on lamp elevation

The classical circumzenith arc occurs for sun elevations less than 32°, and the classical circumhorizon arc occurs for sun elevations greater than 58°. Equivalently: the streetlight circumzenith arc is eye connected for lamp elevations less than 32°, and the streetlight circumhorizon arc is eye-connected for lamp elevations greater than 58°. But both surfaces occur together, at all positive lamp elevations. Figures 7.3–7.7 show the circumzenith and circumhorizon arcs for several lamp elevations.

I do not want to get tangled up here in the daunting problem of halo intensity for spatial halos, but I will say that halo surfaces are better seen when the observer is looking nearly tangentially along them. So in Fig. 7.4, for example, the circumzenith arc should show much better than the circumhorizon arc.

7.3 θ-surface

For a tangent arc—the entire tangent arc, not just the caustic—the points where the axes of the halo-making wedges have azimuth θ make up not a

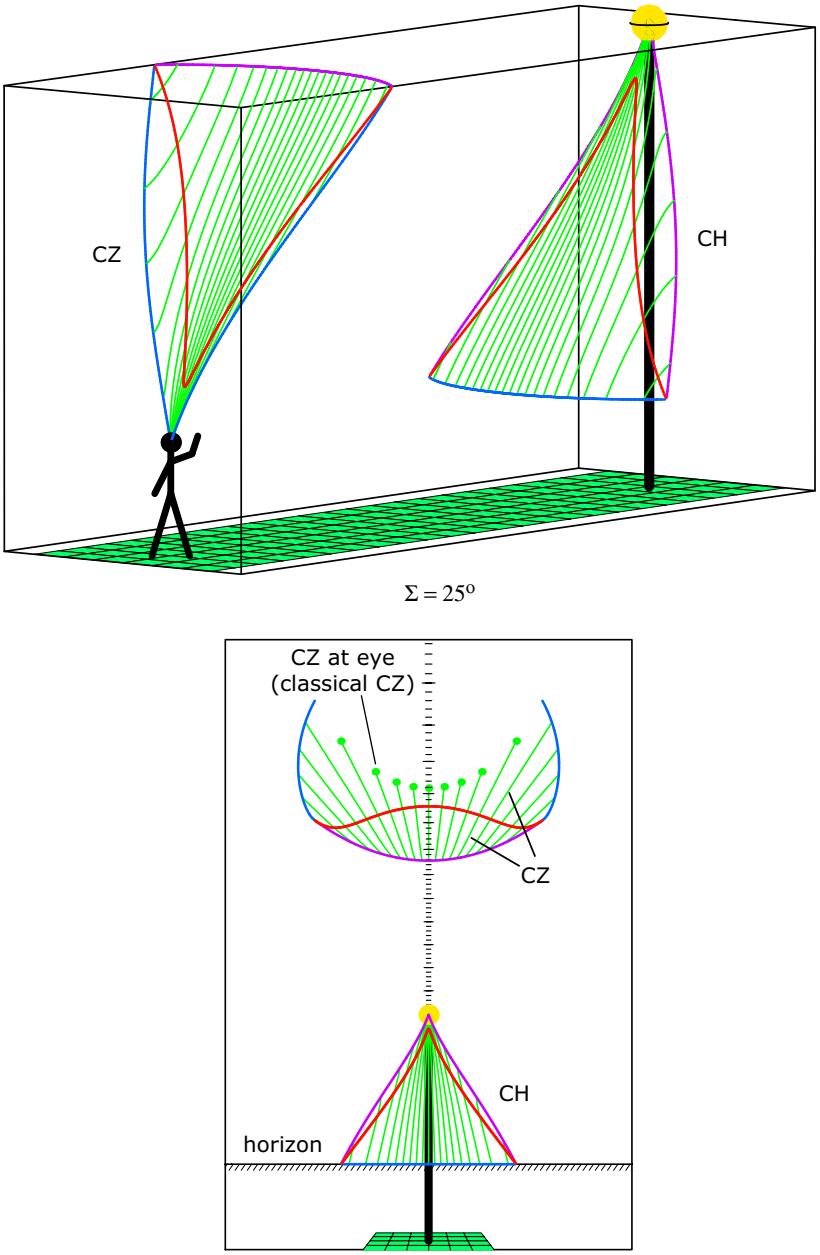

Figure 7.4: (*Top*) Circumzenith and circumhorizon arcs, the same as in Fig. 7.3 but for lamp elevation $\Sigma = 25°$ instead of 15°. (*Bottom*) Observer's view of the arcs. The green dots trace out the classical circumzenith arc. The circumhorizon arc is not eye-connected here, hence the classical circumhorizon arc is empty.

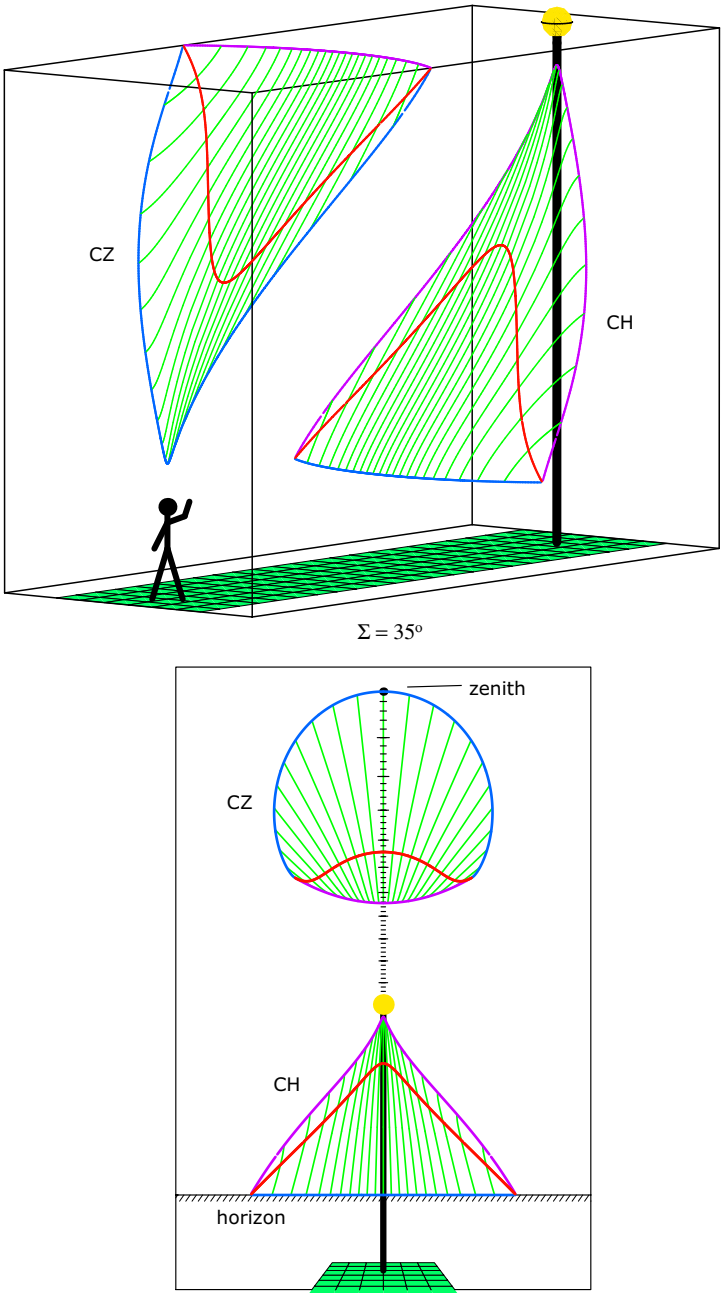

Figure 7.5: (*Top*) Circumzenith and circumhorizon arcs for $\Sigma = 35°$. (*Bottom*) Observer's view of them. At this lamp elevation, neither arc is eye-connected, and so there is no classical instance of either. The lens focal length is 14 mm.

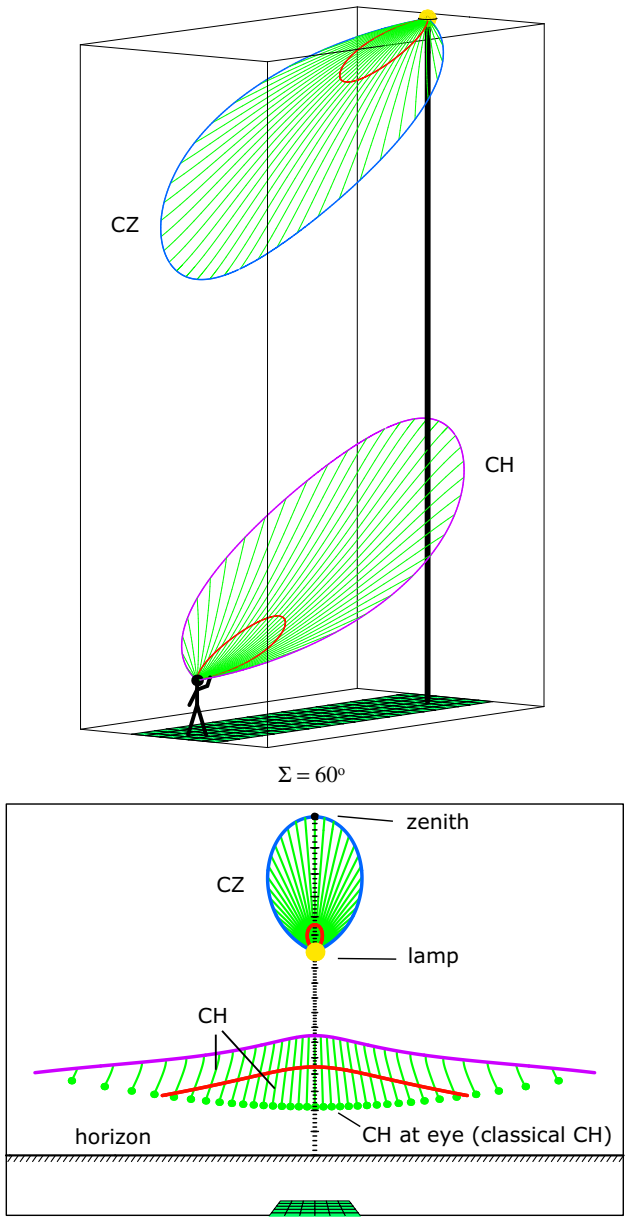

Figure 7.6: (*Top*) Circumzenith and circumhorizon arcs for $\Sigma = 60°$. The circumhorizon arc is now eye-connected. (*Bottom*) Observer's view of the arcs. Because the circumhorizon arc is eye-connected, its classical instance, traced out by the green dots, is non-empty. The lamp pole has been omitted in order to show the tick marks. The lens focal length is 9.5 mm.

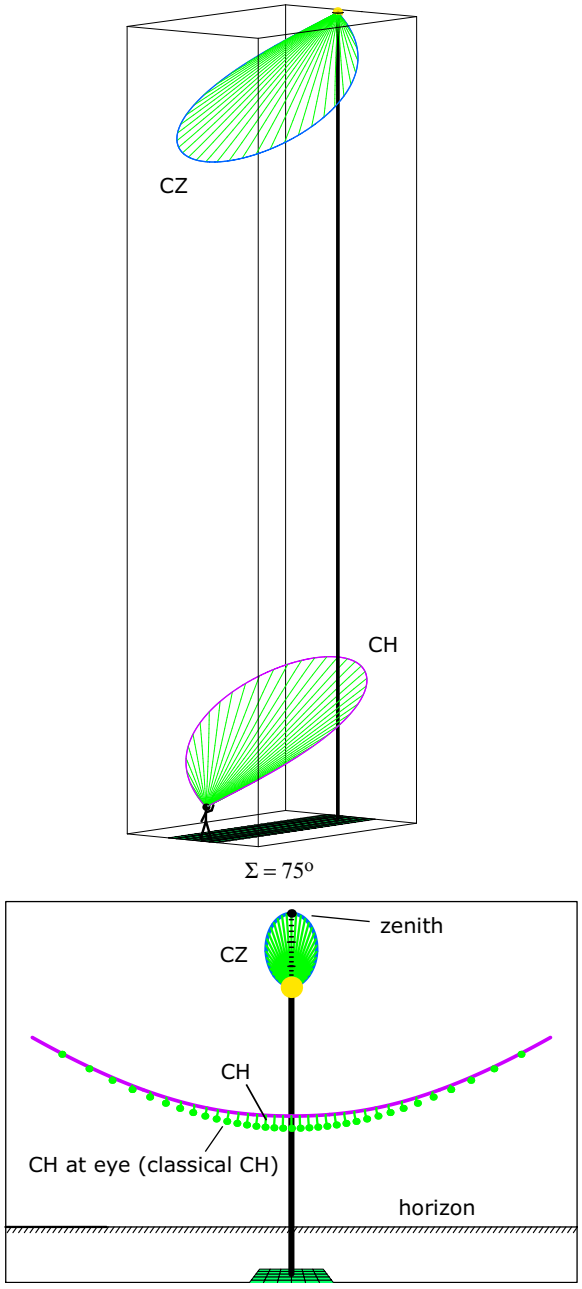

Figure 7.7: (*Top*) Circumzenith and circumhorizon arcs for $\Sigma = 75°$. (*Bottom*) Observer's view of them. The green dots trace out the classical circumhorizon arc. The lens focal length is 9.5 mm.

curve but rather a surface, the 'θ-surface' of the halo. Since the circumzenith arc is a subset of the 46° tanarc, the θ-contour of the circumzenith arc is on the θ-surface of the 46° tanarc. Both are shown in Fig. 7.8 for $\theta = -20°$ (and $\Sigma = 25°$).

The θ-surface of the 46° tanarc should look familiar from Chapter 3, since it is (congruent to) a 46° parhelion; that was Huygens' insight, namely, that the tanarc is made up of parhelia (Section 5.5). Figures 7.9 and 7.10 are reminders of the sorts of halo-making wedges that you have at each point of the θ-surface. Then Fig. 7.11 illustrates how you construct the θ-contour of the circumzenith arc; from the θ-surface you just select the points where the entry faces of the halo-making wedges of the 46° tanarc are horizontal. Remember that in all of Figs. 7.9–7.11 the view is looking directly along the wedge axes, that is, looking in the horizontal direction with azimuth $\theta = -20°$. So the plane of the paper is a common normal plane for the wedges, but nothing in the figures—wedges, curves, eye, lamp—is apt to be in that plane. Thus, for example, the lamp elevation at first glance looks too big—larger than 25°—especially in Fig. 7.12, but that is because the eye and the lamp cannot both be in the plane of the paper.

Figure 7.8 showed how the $\theta = -20°$ contour of the circumzenith arc was constructed. Figure 7.12 does the same for the $\theta = -40°$ contour. Repeating the construction for the other values of θ produces the circumzenith arc—the upper diagram in the two figures.

Similarly, if you wanted to make the 46° tanarc caustic, you would take the minimum deviation curves (red) from θ-surfaces like those in Fig. 7.8 and 7.12, but for many different θ's. And if you wanted to make the entire 46° tanarc—a three-dimensional region—then you would take the entire θ-surface for many θ's. Of course in all of these constructions you have to keep track of how to get the normal plane views back into space.

The classical circumzenith arc is perhaps the simplest refraction halo. Who would have thought that the streetlight version would be so interesting?

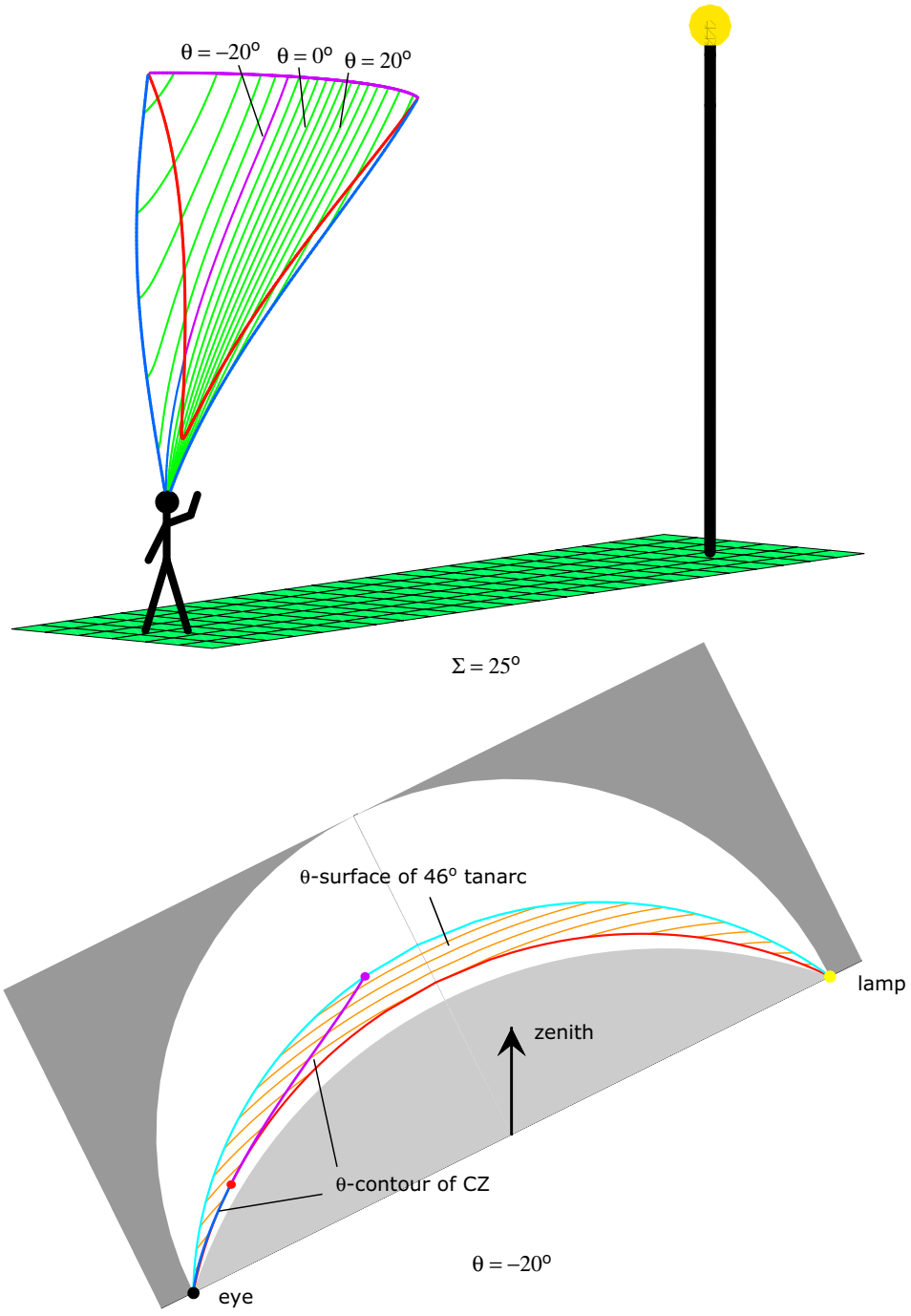

Figure 7.8: (*Top*) Circumzenith arc for $\Sigma = 25°$, the same as in Fig. 7.4. Three θ-contours are labelled, and the $\theta = -20°$ contour has been recolored purple and blue. (*Bottom*) The same $\theta = -20°$ contour together with the $\theta = -20°$ surface of the 46° tanarc. The latter is (congruent to) a 46° parhelion. The view is in the horizontal direction with azimuth $\theta = -20°$.

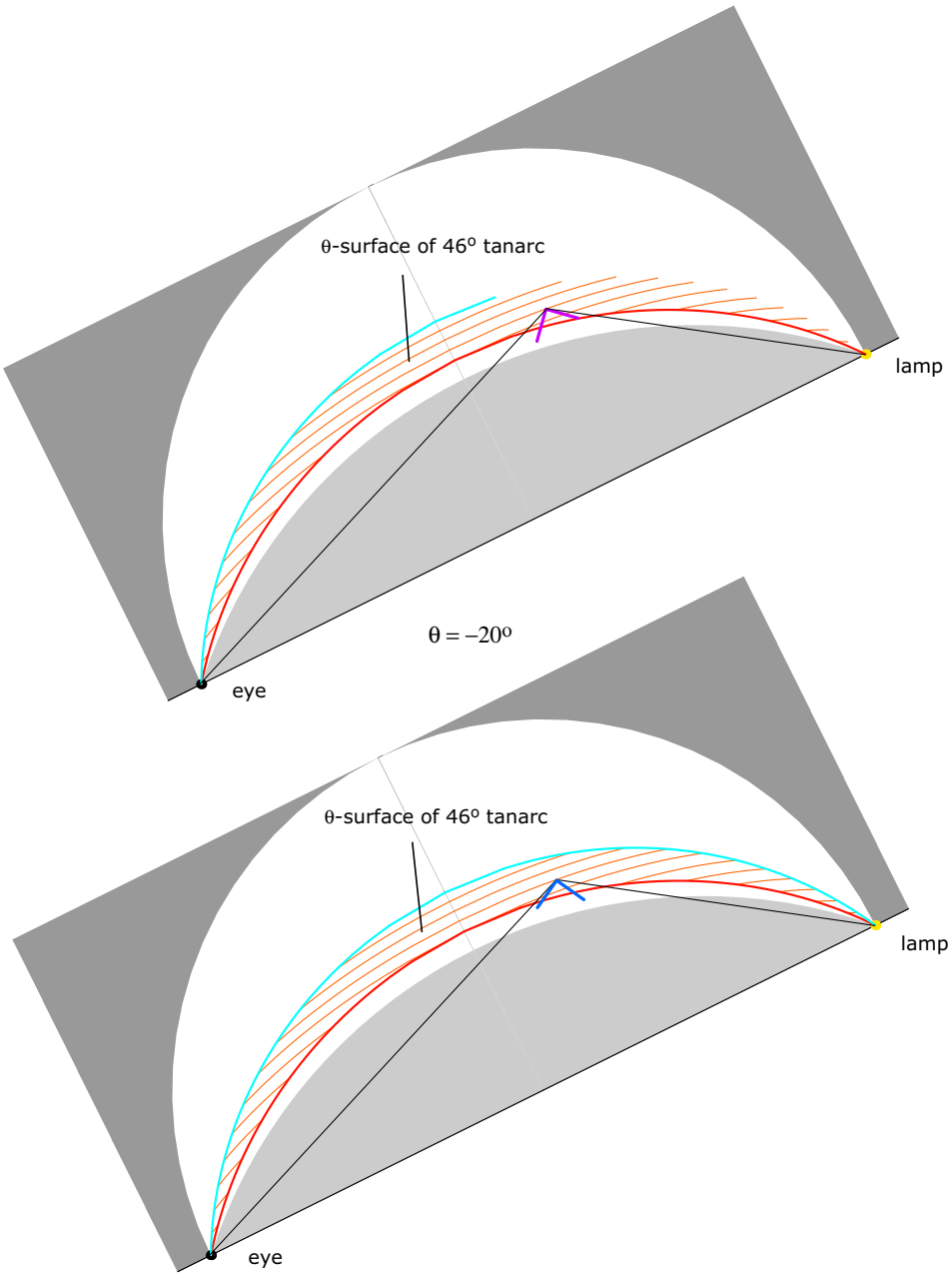

Figure 7.9: The two wedges, at a typical point on the $\theta = -20°$ surface of the 46° tanarc, that refract lamp light to the eye. The ray entry angle for the one wedge is the ray exit angle for the other. For the purple wedge, the entry ray is more nearly tangential than the exit ray. For the blue wedge, the exit ray is more nearly tangential then the entry. ($\Sigma = 25°$.)

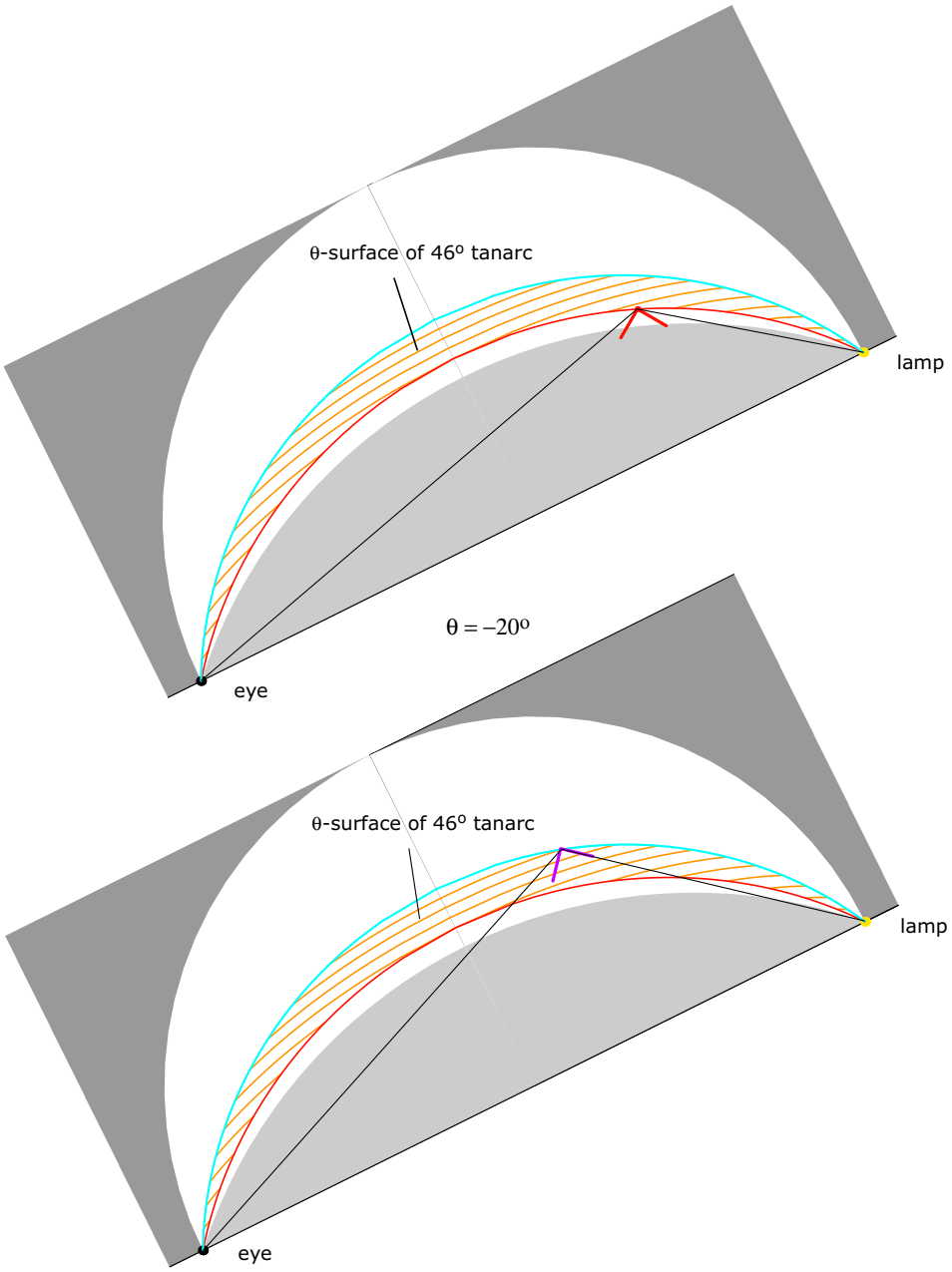

Figure 7.10: Similar to Fig. 7.9 but showing two more wedges, both more special. For the red wedge, on the minimum deviation boundary curve of the θ-surface, the entry and exit angles are equal, and there is just one wedge at this point that refracts lamp light to the eye. For the purple wedge, on the maximum deviation boundary curve, the entry ray is tangent to the wedge.

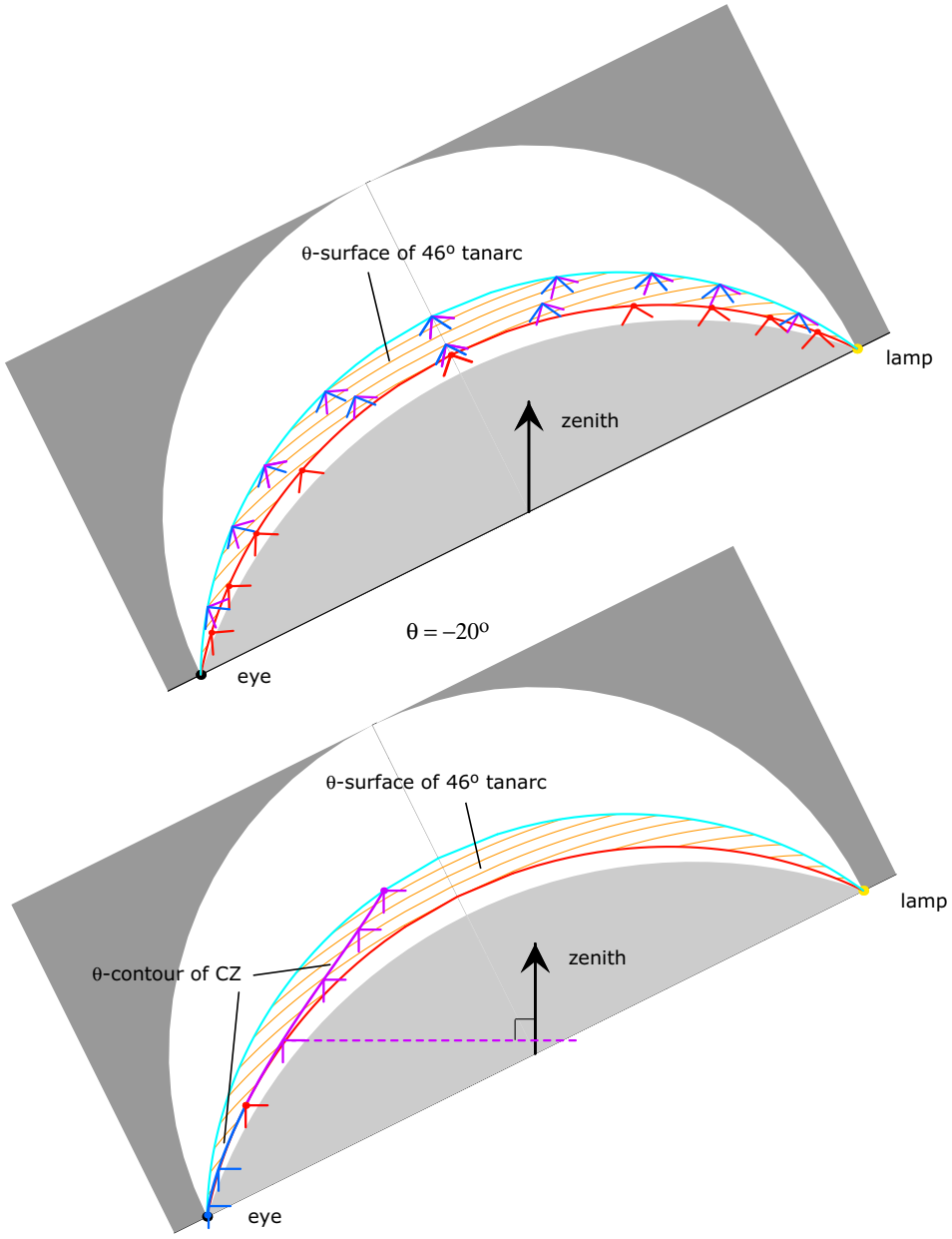

Figure 7.11: (*Top*) The θ-surface of the 46° tanarc as in Fig. 7.9 and Fig. 7.10, but now showing more of the halo-making wedges. Most of the wedges are not contributing to the circumzenith arc. (*Bottom*) Finding the θ-contour (purple and blue) of the circumzenith arc. It consists of the points on the θ-surface of the 46° tanarc where the entry faces of the halo-making wedges are horizontal. This diagram is the same as the bottom diagram in Fig. 7.8 but with the addition of the wedges.

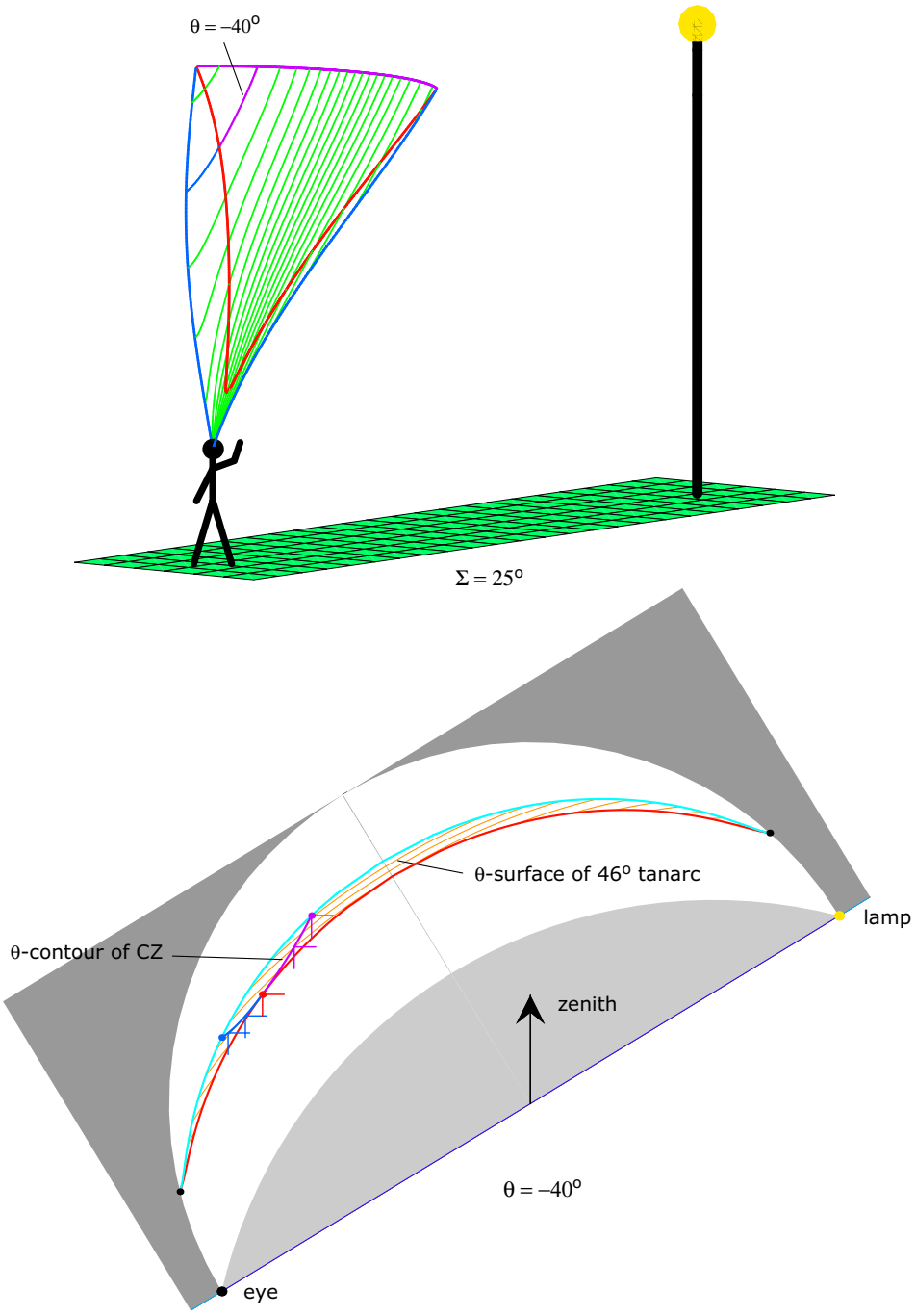

Figure 7.12: Like Fig. 7.8 but with $\theta = -40°$ instead of $\theta = -20°$.